



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁴ : C12N 15/52, 15/67, 15/80 C12N 15/28, 15/26	A1	(11) International Publication Number: WO 86/ 06097 (43) International Publication Date: 23 October 1986 (23.10.86)
(21) International Application Number: PCT/GB86/00209 (22) International Filing Date: 14 April 1986 (14.04.86) (31) Priority Application Numbers: 479,135 811,404 (32) Priority Dates: 15 April 1985 (15.04.85) 20 December 1985 (20.12.85) (33) Priority Countries: CA US (71) Applicant: ALLELIX, INC. [CA/CA]; 6850 Goreway Drive, Mississauga, Ontario (CA). (72) Inventors: GWYNNE, David, Ivor ; 34 Monterrey Avenue, Brampton, Ontario (CA). BUXTON, Francis, Paul ; 727 Indian Road, Toronto, Ontario (CA). PICKETT, Mark, Hudson ; 3 Knightsbridge Road, Apt. 1105, Brampton, Ontario (CA). DAVIES, Roger, Wayne ; R. R. # 1, Limehouse, Ontario (CA). SCAZZOCHIO, Claudio ; 6 parc de Bures, 14, rue de Royaume, F-91440 Bures sur Yvette (FR).		(74) Agent: EYLES, Christopher, Thomas; Batchellor, Kirk & Eyles, 2 Pear Tree Court, Farringdon Road, London EC1R ODS (GB). (81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), DK, FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: PROMOTERS OF FILAMENTOUS FUNGI AND USE THEREOF (57) Abstract <p>Novel vectors for use in filamentous fungi such as <i>Aspergillus sp.</i> in particular, whereby protein coding regions may be inserted therein to achieve expression or expression followed by secretion of the coded protein from the host. Signal peptide sequences and promoter sequences valuable for this purpose are disclosed as are expression vectors containing coding regions native or foreign to the fungal host. In accordance with the invention, a filamentous fungus such as <i>Aspergillus</i> may be provided with foreign or natural coding regions associated with foreign or natural promoter sequences and optionally signal peptide sequences which can be used to control the expression and/or secretion of the proteins encoded by these coding regions.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GA	Gabon	MR	Mauritania
AU	Australia	GB	United Kingdom	MW	Malawi
BB	Barbados	HU	Hungary	NL	Netherlands
BE	Belgium	IT	Italy	NO	Norway
BG	Bulgaria	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali		
FR	France				

PROMOTERS OF FILAMENTOUS FUNGI AND USE THEREOF

This invention relates to expression and expression followed by secretion of proteins from filamentous fungi.

BACKGROUND OF THE INVENTION

One goal of recombinant DNA technology is the insertion of DNA segments which encode commercially or scientifically valuable proteins into a host cell which is readily and economically available. Genes selected for insertion are normally those which encode proteins produced in only limited amounts by their natural hosts or those which are indigenous to hosts too costly to maintain. Transfer of the genetic information in a controlled manner to a host which is capable of producing the protein in either greater yield or more economically in a similar yield provides a more desirable vehicle for protein production.

Genes encoding proteins contain promoter regions of DNA which are essentially attached to the 5' terminus of the protein coding region. The promoter regions contain the binding site for RNA polymerase II. RNA polymerase II effectively catalyses the assembly of the messenger RNA complementary to the appropriate DNA strand of the coding region. In most promoter regions, a nucleotide base sequence related to the sequence known generally as a "TATA box" is present and is generally disposed some distance upstream from the start of the coding region and is required for accurate initiation of transcription. Other features important or essential to the proper functioning and control of the coding region are also contained in the promoter region, upstream of the start of the coding region.

Filamentous fungi, particularly the filamentous ascomycetes such as Aspergillus, e.g. Aspergillus niger, represent a class of micro-organisms suitable as recipients of

- 2 -

foreign genes coding for valuable proteins. Aspergillus niger and related species are currently used widely in the industrial production of enzymes e.g. for use in the food industry. Their use is based on the secretory capacity of the microorganism. Because they are well characterized and because of their wide use and acceptance, there is both industrial and scientific incentive to provide genetically modified and enhanced cells of A. niger and related species including A. nidulans, in order to obtain useful proteins.

Expression and secretion of foreign proteins from filamentous fungi has not yet been achieved. It is by no means clear that the strategies which have been successful in yeast would be successful in filamentous fungi such as Aspergillus. Evidence has shown that yeast is an unsuitable system for the expression of filamentous fungal genes (Pentilla et al Molec. Gen. Genet. (1984) 194:494-499) and that yeast genes do not express in filamentous fungi. Genetic engineering techniques have only recently been developed for Aspergillus nidulans and Aspergillus niger. These techniques involve the incorporation of exogenously added genes into the Aspergillus genome in a form in which they are able to be expressed.

To date no foreign proteins have been expressed in and secreted from filamentous fungi using these techniques. This has been due to a lack of suitable expression vectors and their constituent components. These components include Aspergillus promoter sequences described above, the region encoding the desired product and the associated sequences which may be added to direct the desired product to the extracellular medium.

As noted, expression of the foreign gene by the host cell requires the presence of a promoter region situated upstream of the region coding for the protein. This promoter region is active in controlling transcription of the coding region with which it is associated, into messenger RNA which is ultimately translated into the desired protein product.

- 3 -

Proteins so produced may be categorized into two classes on the basis of their destiny with respect to the host.

A first class of proteins is retained intracellularly. Extraction of the desired protein, when intracellular, requires
5 that the genetically engineered host be broken open or lysed in order to free the product for eventual purification.

Intracellular production has several advantages. The protein
product can be concentrated i.e. pelleted with the cellular
mass, and if the product is labile under extracellular
10 conditions or structurally unable to be secreted, this is a desired method of production and purification.

A second class of proteins are those which are secreted from the cell. In this case, purification is effected on the extracellular medium rather than on the cell itself. The
15 product can be extracted using methods such as affinity chromatography and continuous flow fermentation is possible. Also, certain products are more stable extracellularly and are benefited by extracellular purification. Experimental evidence has shown that secretion of proteins in eukaryotes is almost
20 always dictated by a secretion signal peptide (hereafter called signal peptide) which is usually located at the amino terminus of the protein. Signal peptides have characteristic distributions as described by G. Von Heijne in Eur. J. Biochem 17-21 (1983) and are recognizable by those skilled in the art.
25 The signal peptide, when recognized by the cell, directs the protein into the cell's secretory pathway. During secretion, the signal peptide is cleaved off making the protein available for harvesting in its mature form from the extracellular medium.

Both classes of protein, intracellular and
30 extracellular, are encoded by genes which contain a promoter region coupled to a coding region. Genes encoding extracellularly directed proteins differ from those encoding intracellular proteins in that, in genes encoding extracellular proteins, the portion of the coding region nearest to the

- 4 -

promoter (which is the first part to be transcribed by RNA polymerase) encodes a signal peptide. The nucleotide sequence encoding the signal peptide, hereafter denoted the signal peptide coding region or the signal sequence, is operationally part of the coding region per se.

SUMMARY OF THE INVENTION

A system has now been developed by which filamentous fungi may be transformed to express a desired protein. With this system, transformation can result in a filamentous fungus which is capable not only of expressing the protein but of secreting that protein as well, regardless of whether or not the protein is a naturally secreted one. In addition, the level at which the protein is expressed can be controlled according to certain aspects of the invention. It will be appreciated by those skilled in the art that the system provided hereby permits filamentous fungi to function as valuable sources of proteins and provides an alternative which in many applications is superior to bacterial and yeast systems.

Thus, in a general aspect, the invention provides a filamentous fungus transformation system by which the genetic constitution of these fungus cells may be modified so as to alter either the nature or the amount of the proteins expressed by these cells. More specific aspects of the invention are defined below.

In the present invention, from one aspect, a promoter region associated with a coding region in filamentous fungi such as A. niger, A. nidulans or a related species is identified and isolated, appropriately joined in a functional relationship with a second, different coding region, outside the cell, and then re-introduced into a host filamentous fungus using an appropriate vector. Transformed host cells express the protein of the second coding region, under the control of the introduced promoter region. The second coding region may be one which is

- 5 -

foreign to the host species, in which case the host will express and in some cases secrete a protein not naturally expressed by the given host. Alternatively, the second coding region may be one which is natural to the host, in which case it is associated with a promoter region different from the promoter region with which it naturally associates in the given host, to give modified or enhanced protein expression and secretion.

Where the second coding region is one which encodes a protein which is normally secreted, it will contain a sequence of nucleotides at its 5' terminus i.e. a signal peptide coding region, which will result, following transcription and translation, in the presence of a signal peptide at the amino terminus of the protein product. The signal peptide can be recognized by the fungal host and the protein product can then be directed into the secretory pathway of the cell and secreted.

In another aspect, the present invention provides DNA sequences coding for a signal peptide i.e. a signal peptide coding region, which is recognized by filamentous fungi preferably of the ascomycetes class e.g. A. niger and A. nidulans, so as to signal secretion of a protein encoded within the coding region. These signal peptide coding regions can be coupled to a coding region which encodes a protein naturally retained intracellularly in order to elicit secretion of that protein. While normally secreted proteins are encoded by coding regions which usually contain these signal peptide coding regions naturally so that incorporation of a signal peptide coding region is not usually necessary, the signal peptide coding regions of the present invention may nevertheless be substituted for the naturally occurring such sequence, if desired. Accordingly, where a signal peptide coding region is coupled to a region encoding a non-secreted protein, it will be foreign to that coding region.

The present invention provides the ability to introduce foreign coding regions into filamentous fungi along with

- 6 -

promoters to arrange for the host fungi to express different proteins. It also provides the ability to regulate transcription of the individual genes which occur naturally therein or foreign genes introduced therein, via the promoter region which has been introduced into the host along with the gene. For example, the promoter region naturally associated with the alcohol dehydrogenase I (alca) gene and the aldehyde dehydrogenase (aldA) gene of A. nidulans are regulatable by means of ethanol, threonine, or other inducing substances in the extracellular medium. This effect is dependent on the integrity of a gene known as alcR. When the alca or aldA promoter region is associated with a foreign protein coding region in Aspergillus or the like, in accordance with the present invention, similar regulation of the expression of the different genes by ethanol or other inducers can be achieved.

As a further example, the promoter region naturally associated with the glucoamylase gene in Aspergillus niger and used in embodiments of the present invention is positively induced with starch and other sugars.

In another aspect, the present invention provides a DNA construct which contains a promoter region in operative association with a signal peptide coding region and which permits introduction of a region coding for a desired protein at a position 3' of and in reading frame with the signal peptide coding region. The promoter/signal construct is suitably provided with a flanking restriction site to allow precise coupling of the protein coding region to the signal peptide coding region.

In another aspect, the present invention provides a genetic vector capable of introducing the segment carrying the promoter and signal peptide coding region with integral protein coding region into the genome of a filamentous fungus host. The protein coding region can be either native to or foreign to the host filamentous fungus.

Thus the present invention, provides DNA sequences active as promoter regions and DNA sequences active as signal peptide coding regions in cells of filamentous fungi such as Aspergillus niger, Aspergillus nidulans and the like.

5 The present invention thus also provides a novel construct comprising a DNA sequence active as a promoter region in cells of filamentous fungi, and a coding region chemically bound to said DNA sequence in operative association therewith, said coding region being capable of expression in a filamentous
10 fungus host under influence of said DNA sequence.

 The present invention further provides a process of genetically modifying a filamentous fungus host cell which comprises introducing into the host cell, by means of a suitable vector, a coding region capable of expression in the transformed
15 Aspergillus host cell and a promoter region active in the transformed Aspergillus host cell, the coding region and the promoter being chemically bound together and in operative association with one another.

 This process also encompasses the introduction of
20 multiple copies of the selected construct into the host to provide for enhanced levels of gene expression. If necessary or desirable, introduction of multiple construct copies is accompanied by introduction of multiple copies of genes encoding products having a regulatory effect on the construct.

25 The present invention also comprises filamentous fungal cells transformed by the constructs of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

 Preferred hosts according to the invention are the filamentous fungi of the ascomycete class, most preferably
30 Aspergillus sp. including A. niger, A. nidulans and the like.

- 8 -

In the preferred form of the invention the promoter region associated with either the Aspergillus niger glucoamylase gene or the promoter region associated with either the alcohol dehydrogenase I gene or aldehyde dehydrogenase genes of Aspergillus nidulans is used in preparing an appropriate vector plasmid.

Either or all of these promoter regions is regulatable in the host cell by the addition of the appropriate inducer substance. In alcA and aldA, this induction is mediated by the protein product of a third gene, alcR which is controlled via the promoter. Evidence indicates that the availability of alcR product can limit the promoting function of the alcA and aldA promoters when multiple copies of a construct containing the alcA promoter or the aldA promoter are introduced into a host without corresponding introduction of multiple copies of the alcR gene. In such a case, the amount of alcR product which the host can produce may be insufficient to meet the demands of the several promoters requiring induction by the alcR product. Thus, transformation of filamentous fungal hosts by multiple copies of constructs containing the alcA or aldA promoter is accompanied by introduction of multiple copies of the alcR gene, according to a preferred embodiment of the present invention. In other instances, transcription can be repressed, for example by utilizing high levels of glucose, (and some other carbon sources) in the medium to be used for growth of the host. The expression of the product encoded by the coding region and controlled by the promoter is then delayed until after the end of the cell growth phase, when all of the glucose has been consumed and the gene is derepressed. The inducer may be added at this point to enhance the activity of the promoter.

The destination of the protein product of the coding region which has been selected to be expressed under the control of the promoter described above is determined by the nucleotide sequence of that coding region. As mentioned, if the protein product is naturally directed to the extracellular environment,

- 9 -

it will inherently contain a secretion signal peptide coding region. Protein products which are normally intracellularly located lack this signal peptide.

Thus, for the purposes of the present disclosure it is to be understood that a "coding region" encodes a protein which is either retained intracellularly or is secreted. (This "coding region" is sometimes referred to in the art as a structural gene i.e. that portion of a gene which encodes a protein.) Where the protein is retained within the cell that produces it, the coding region will usually lack a signal peptide coding region. Secretion of the protein encoded within the coding region can be a natural consequence of cell metabolism in which case the coding region inherently contains a signal peptide coding region linked naturally in translation reading frame with that segment of the coding region which encodes the secreted protein. In this case, insertion of a signal peptide coding region is not required. In the alternative, the coding region may be manipulated to introduce a signal peptide coding region which is foreign to that portion of the coding region which encodes the secreted protein. This foreign signal peptide coding region may be required where the coding region does not naturally contain a signal peptide coding region or it may simply replace the natural signal peptide coding region in order to obtain enhanced secretion of the desired protein with which the natural signal peptide is normally associated.

In accordance with another preferred aspect of the invention, therefore, a signal peptide coding region is provided, if required i.e. when the coding region which has been selected to be expressed under the control of the promoter described above does not itself contain a signal peptide coding region. The signal peptide coding region used is preferably either one which is associated with the Aspergillus niger glucoamylase gene or a synthetic signal peptide coding region which is made in vitro and used in the preparation of an

- 10 -

appropriate vector plasmid. Most preferably, these signal peptide coding regions are modified at one or both termini to permit ligation thereof with other components of a vector. This ligation is effected in such a way that the signal peptide coding region is interposed between the promoter region and the protein encoding segment of the coding region such that the signal peptide coding region is in frame with that segment of the coding region which encodes the mature, functional protein.

BRIEF REFERENCE TO THE DRAWINGS

Figure 1A is an illustration of the base sequence of the DNA constituting the coding region and promoter region of the alcohol dehydrogenase I (alca) gene of Aspergillus nidulans.

Figure 1B is an illustration of the base sequence of DNA constituting the coding region and promoter region of the aldehyde dehydrogenase (aldA) gene of Aspergillus nidulans.

Figure 2 is a diagrammatic illustration of a process of constructing plasmid pDG6 useful in transforming a filamentous fungal cell;

Figure 3 is a linear representation of a portion of the plasmid pDG6 of Fig.2;

Figure 4 is a diagrammatic illustration of the plasmid maps of pGL1 and pGL2;

Figure 5 is an illustration of a selection of synthetic linker sequences for insertion into plasmid pGL2;

Figure 6 is an illustration of the nucleotide sequence of a fragment of pGL2;

Figure 7 is an illustration of plasmid map pGL2B and pGL2BIFN;

- 11 -

Figure 8 is an illustration of the nucleotide sequence of a fragment of pGL2BIFN;

Figure 9 illustrates plasmid pALCA1S and a method for its preparation;

5 Figure 10 illustrates the plasmid map of pALCA1SIFN and a method for its preparation;

Figure 11 represents the nucleotide sequence of a fragment of pALCA1SIFN;

Figure 12 illustrates the plasmid map of pGL2CENDO;

10 Figure 13 represents the nucleotide sequence of a fragment of pGL2CENDO;

Figure 14 represents a plasmid map of pALCA1SENDO;

Figure 15 represents the nucleotide sequence of a fragment of pALCA1SENDO;

15 Figure 16 illustrates plasmid pALCA1AMY and a method for its preparation; and

Figure 17 represents the nucleotide sequence of a segment of pALCA1AMY shown in Figure 16.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

20 In the present invention, an appropriate promoter region of a functioning gene in A. niger or A. nidulans or the like is identified. Procedures for identifying each of the genes containing the desired promoter regions are similar and for that reason, the manner of locating and identifying the alca
25 gene and promoter therein is outlined. For this purpose, cells of the chosen species are induced to express the selected

- 12 -

protein e.g. alca, and from these cells is isolated the messenger RNA. One portion thereof, as yet unidentified codes for alca. Complementary DNA for the fragments is prepared from the mRNA fragments and cloned into a vector. Messenger RNA
5 isolated from induced A. nidulans is size fractionated to enrich for alca sequences, end labelled and hybridized to the cDNA clones made from the alca⁺ strain. That clone containing the cDNA which hybridizes to alca⁺ mRNA contains the DNA copy of the alca mRNA. This piece is hybridized to a total DNA gene
10 bank from the chosen Aspergillus species, to isolate the selected coding region e.g. alca and its flanking regions. The aldA coding region was isolated using analogous procedures.

The coding region starts at its 5' end, with a codon ATG coding for methionine, in common with other coding regions
15 and proteins. Where the amino acid sequence of the expressed protein is known, the DNA sequence of the coding region is readily recognizable. Immediately "upstream" of the ATG codon is the leader portion of the messenger RNA preceded by the promoter region.

20 With reference to Figs. 1A and 1B, these show portions of the total DNA sequence from A. nidulans, with conventional base notations. The portion shown in Figure 1A contains the promoter region and the coding region of the alca gene which encodes the enzyme alcohol dehydrogenase I. The portion shown
25 in the Figure 1B contains the promoter region and the region encoding the enzyme aldehyde dehydrogenase i.e. aldA. (In both cases, the term "IVS" represents intervening sequences.) The amino acid sequences of these two enzymes is known in other species. From these, the regions 10 and 10' are recognisable as
30 the coding regions. Each coding region starts at its 5' ("upstream") end with methionine codon ATG at 12. The appropriate amino acid sequences encoded by the protein coding region are entered below the respective rows on Figs. 1A and 1B, in conventional abbreviations. Immediately upstream of codon 12
35 is the region coding for the messenger RNA leader and the

- 13 -

promoter region, the length of which, in order to contain all the essential structural features enabling it to function as a promoter, now needs to be determined or at least estimated. Each of Figures 1A and 1B shows a sequence of about 800 bases in each case, upstream from the ATG codon 12.

It is predictable from analogy with other known promoters that all the functional essentials are likely to be contained within a sequence of about 1000 bases in length, probably within the 800 base sequence illustrated, and most likely within the first 200 - 300 base sequence, i.e. back to about position 14 on Figs. 1A and 1B. An essential function of a promoter region is to provide a site for accurate initiation of transcription, which is known to be a TATA box sequence. Such a sequence is found at 16 on the alca promoter sequence of Fig. 1A, and at 16' on the aldA promoter sequence of Fig. 1B. Another function of a promoter region is to provide an appropriate DNA sequence active in regulation of the gene transcription, e.g. a binding site for a regulatory molecule which enhances gene transcription, or for rendering the gene active or inactive. Such regulator regions are within the promoter region illustrated in Figs. 1A and 1B for the alca and aldA genes, respectively.

The precise upstream 5' terminus of the DNA sequence used herein as a promoter region is not critical, provided that it includes the essential functional sequences as described herein. Excess DNA sequences upstream of the 5' terminus are unnecessary, but unlikely to be harmful in the present invention.

Having determined the extent of the sequence containing all the essential functional features to constitute a promoter region of the given gene, by techniques described herein, the next step is to cut the DNA chain at a convenient location downstream of the promoter region terminus and to remove the protein coding region, to leave basically a sequence comprising the promoter region and sometimes part of the region coding for

- 14 -

the messenger RNA leader. For this purpose, appropriately positioned restriction sites are to be located, and then the DNA treated with the appropriate restriction enzymes to effect scission. Restriction sites are recognizable from the alca sequence illustrated in Figure 1A. For the upstream cutting, a site is chosen sufficiently far upstream to include in the retained portion all of the essential functional sites for the promoter region. As regards the downstream scission, no restriction site presents itself exactly at the ATG codon 12 in the case of alca. The closest downstream restriction site thereto is the sequence GGGCCC at 13, at which the chain can be cut with restriction enzyme Apa I. If desired, after such scission, the remaining nucleotides from location 13 to location 12 can be removed, in stepwise fashion, using an exonuclease. With knowledge of the number of such nucleotides to be removed, the exonuclease action can be appropriately stopped when the location 12 is passed. By locating a similar restriction site downstream of the methionine codon 12 of the aldA coding region shown in Fig. 1B, this promoter region is similarly excised for subsequent use. In many cases, residual nucleotides on the 5' terminus of the promoter region are not harmful to and do not significantly interfere with the functioning of the promoter region, so long as the reading frame of the base triplets is maintained.

Fig. 2 of the accompanying drawings illustrates diagrammatically the steps in a process of preparing plasmid pDG6 which can be used to create Aspergillus transformants according to the present invention. On Fig. 2, 18 is a recombinant plasmid containing the endoglucanase (cellulase) coding region 30 from the bacterium Cellulomonas fimi, namely a BamHI endoglucanase fragment from C. fimi in known vector M13MP8. It contains relevant restriction sites for EcoRI, Hind III and BamHI as shown as well as others not shown and not of consequence in the present process. Item 20 is a recombinant plasmid designated p5, constructed from known E. coli plasmid pBR322 and containing an EcoRI fragment of A. nidulans

containing the alcA promoter region prepared as described above, along with a small portion of the alcA coding region, including the start codon ATG. It has restriction sites as illustrated, as well as other restriction sites not used in the present process and so not illustrated. Plasmid p5 contains a DNA sequence 22, from site EcoRI (3') to site Hind III (5'), which is in fact a part of the sequence illustrated on Fig. 1A, upper row, from position 15 (the sequence GAATTC thereat constituting an EcoRI restriction site) to position 17. Sequence 22 in plasmid 20 is approximately 2 kb in length.

The plasmids 18 and 20 are next cut with restriction enzymes EcoRI and Hind III, so as to excise the alcA promoter region and the endoglucanase coding region 30 which are ligated to Hind III-cut plasmid pUC12, to form a novel construct pDG5A containing these sequences on pUC12, as shown in Fig. 2. Plasmid pUC12 is a known, commercially available E. coli plasmid, which replicates efficiently in E. coli, so that abundant copies of pDG5A can be made if desired. Novel construct pDG5A is isolated from the other products of the construct preparation. Next, construct pDG5A is provided with a selectable marker so that subsequently obtained transformants of Aspergillus into which the construct has successfully entered can be selected and isolated. In the case of Arg B⁻ Aspergillus hosts, one can suitably use an Arg B gene from A. nidulans for this purpose. The Arg B gene codes for the enzyme ornithine transcarbamylase, and strains containing this gene are readily selectable and isolatable from Arg B⁻ strains by standard plating out and cultivation techniques. Arg B⁻ strains will not grow on a medium lacking arginine.

To incorporate a selectable marker, in this embodiment of the invention as illustrated in Fig. 2, construct pDG5A may be ligated with the Xba I fragment 32 of plasmid pDG3 ATCC 53006 (see U.S. patent application serial number 06/678,578 Buxton et al, filed December 5, 1984) which contains the Arg B⁺ gene from A. nidulans using Xba I, to form novel construct pDG6,

which contains the endoglucanase coding region, the alca promoter sequence and the Arg B gene. Plasmid pDG6 is then used in transformation, to prepare novel Aspergillus mutant strains containing an endoglucanase coding region under the control of alca promoter, as described in more detail in Example 1.

Fig. 3 shows in linear form the diagrammatic sequence of the functional portion of construct pDG6, from the Hind III site 24 to the Hind III site 26. It contains the alca promoter region 22, the ATG codon 12 and a small residual portion of the alca coding region downstream of the ATG codon as shown in Fig. 1, followed by the cellulase coding region 30 derived from plasmid 18.

Plasmid pDG6 is but one example of a vector which contains a filamentous fungal promoter linked to a protein coding region foreign to the fungus. In another vector exemplified herein with reference to Figure 16 and referred to herein as pALCALAMY, the filamentous fungal promoter of the alca gene is coupled with the naturally occurring sequence coding for the α -amylase enzyme, a product which is foreign to the transformed fungal host. The protein products of vectors pDG6 and pALCALAMY are expressed by the respective transformed hosts in both cases. Further, because the α -amylase coding region naturally contains a signal peptide coding region, this product can be secreted by the transformed host, using the secretory machinery of the host, despite its foreign relationship with the host.

Identifying and isolating the promoter regions of filamentous fungi thus allows one to manipulate the host by transformation with vectors containing these promoter regions coupled with a desired coding region.

If the coding region of the vector requires a signal peptide coding region or the existing signal sequence is to be replaced by a different, preferably more efficient signal

- 17 -

peptide coding region, such signal peptide coding regions may be integrated between the promoter and that segment coding for the secreted protein. Plasmids pGL2 (Figure 4) and pALCA1S (Figure 9) represent intermediate cloning vectors particularly suited for this purpose. Each can function as a cassette, providing a promoter, a signal sequence and a restriction site downstream of the signal sequence which permits insertion of a protein coding region in proper, transcriptional reading frame with the signal sequence.

Plasmid pGL2 shown in Figure 4 is created from pGL1 which contains the promoter 40, the signal sequence 42 and an initial portion 46 of the glucoamylase gene, all of which were derived in one segment from A. niger DNA according to methods exemplified herein. In this segment, a BssHII restriction site is available toward the end of, but nevertheless within the glucoamylase signal sequence 42, the nucleotide sequence of which is reproduced below in chart 1.

Chart 1

5' ATG TCG TTC CGA TCT CTA CTC GCC CTG AGC GGC CTC GTC TGC
met ser phe arg ser leu leu ala leu ser gly ley val cys

BssH II
↓
ACA GGG TTG GCA AAT GTG ATT TCC AAG CGC 3'
thr gly leu ala asn val ile ser lys arg

In order to provide a segment downstream of the signal sequence i.e. a linker 44, capable of receiving a protein coding region in reading frame with the signal sequence 42, advantage is taken of the presence of the BssH II site within the signal sequence and the Sst I site downstream thereof. In this specific embodiment, segment 46 is excised from pGL1 and replaced with a selected one of three linkers shown in Figure 5 and denoted A, B or C. Each linker is able to ligate with the BssH II end and the Sst I end. The linkers are also engineered

- 18 -

so as to restore the terminal codons of the signal sequence lost upon excision of segment 46 with BssH II. Further, each linker defines unique EcoRV and Bgl II/ Xho II sites within its nucleotide sequence so as to permit insertion of the desired coding region into the vector pGL2.

Selection of the appropriate linker is made with knowledge of at least the first few codons of the protein coding region to be inserted into the linker. In order for the protein coding region to be translated sensibly, the start of the protein coding region must be either directly coupled with or be a specific number of nucleotides i.e. in triplets, from the start of the signal sequence. Accordingly, if the protein coding region to be inserted possesses one or two unessential nucleotides (or a non-triplet factor thereof) at its 5' region as may result from routine excision, one of the three linkers shown in Figure 5 can compensate for the presence of the extra, superfluous nucleotides and locate the start of the protein coding region in translational reading frame with the signal sequence.

The amino acid residues encoded by the linkers A, B and C appear under their nucleotide sequences as shown in Figure 5, from which the effect of adding an additional nucleotide to the linker sequence on the reading frame of the linker and ultimately on the inserted protein encoding region may be noted. By designing the linkers such that the restriction site is always downstream of the reading frame modification i.e. one, two or three adenine residues in linkers A, B or C respectively, the reading frame of the coding region inserted into the restriction site can be maintained by appropriate linker selection.

Exemplary of plasmids which employ plasmid pGL2 and specific linker segments A, B, or C are pGL2BIFN which employs the B linker and results in interferon α -2 secretion when used in filamentous fungus e.g. Aspergillus sp., transformation and

- 19 -

pGL2CENDO which employs the C linker and results in endoglucanase secretion when such filamentous fungi are transformed therewith.

While plasmid pGL2 utilizes a naturally occurring
*5 signal sequence, it is within the scope of the invention also to
utilize vectors containing synthetic signal sequences. An
example of one such vector is pALCA1S which, like plasmid pGL2,
represents an intermediate vector within which a protein coding
region may be inserted to form a vector capable of transforming
10 filamentous fungi. Unlike pGL2 however, pALCA1S utilizes the
alcA promoter and utilizes a synthetic signal sequence coupled
to that promoter. pALCA1S is illustrated in Figure 9 which
shows a scheme for preparing it and to which further reference
is made in the examples. Exemplary of plasmids created from
15 pALCA1S are pALCA1SIFN which results in secretion of
interferon α -2 from a filamentous fungus transformed therewith
and pALCA1SEENDO which results in secretion of endoglucanase from
a filamentous fungus host. In both instances secretion is
obtained despite the foreign nature of the secreted protein with
20 respect to the host.

The invention is further described and illustrated by
the following specific, non-limiting examples.

Each of Examples 1 and 2 which follow exemplify
successful transformation of a filamentous fungal host using
25 vectors having a filamentous fungus-derived promoter coupled
with naturally occurring but non-fungal coding regions.

Example 1 - Transformation of A. nidulans using pDG6 ATCC 53169

The vector construct pDG6 shown in Figure 2 was first
prepared following the process scheme illustrated in Figure 2,
30 using standard routine ligation and restriction techniques.
Then the construct pDG6 was introduced into Arg B⁻ mutant
cells of Aspergillus nidulans as follows:

- 20 -

500 mls of complete media (Cove 1966) + 0.02% arginine + 10^{-5} % biotin in a 2 l conical flask was inoculated with 10^5 conidia/ml of an A. nidulans Arg B⁻ strain and incubated at 30°C, shaking at 250 rpm for 20 hours. The mycelia were harvested through Whatman No. 54 filter paper, washed with sterile deionized water and sucked dry. The mycelia were added to 50 ml of filter sterile 1.2 M MgSO_4 , 10 mM potassium phosphate pH 5.8 in a 250 ml flask to which was added 20 mg of Novozym 234 (Novo Enzyme Industries), 0.1 ml (=15000 units) of β -glucuronidase (Sigma) and 3 mg of Bovine serum albumin for each gram of mycelia. Digestion was allowed to proceed at 37°C with gentle shaking for 50-70 minutes checking periodically for spheroplast production by light-microscope. 50 mls of sterile deionised water was added and the spheroplasts were separated from undigested fragments by filtering through 30 μm nylon mesh and harvested by centrifuging at 2500 g for 5 minutes in a swing out rotor in 50 ml conical bottom tubes, at room temperature. The spheroplasts were washed, by resuspending and centrifuging, twice in 10 mls of 0.6 M KCl. The number of spheroplasts was determined using a hemocytometer and they were resuspended at a final concentration of $10^8/\text{ml}$ in 1.2 M Sorbitol, 10 mM Tris/HCl, 10 mM CaCl_2 pH 7.5. Aliquots of 0.4 ml were placed in plastic tubes to which DNA pDG6 (total vol. 40 μl in 10 mM Tris/HCl 1 mM EDTA pH 8) was added and incubated at room temperature for 25 minutes. 0.4 ml, 0.4 ml then 1.6 ml aliquots of 60% PEG4000, 10 mM Tris/HCl, 10 mM CaCl_2 pH 7.5 were added to each tube sequentially with gentle, but thorough mixing between each addition, followed by a further incubation at room temperature for 20 minutes. The transformed spheroplasts were then added to appropriately supplemented minimal media 1% agar overlays, plus or minus 0.6 M KCl at 45°C and poured immediately onto the identical (but cold) media in plates. After 3-5 days at 37°C the number of colonies growing was counted (F. Buxton et al), Gene 37, 207-214 (1985)). The method of Yelton et al [Proc. Nat'l Acad. Sci. U.S.A. 81; 1370-1374 (1980)] was also used.

- 21 -

The colonies were divided into two groups. Threonine (11.9 g/Liter) and fructose (1 g/Litre) were added to the incubation medium for one group to induce the cellulase gene incorporated therein. No inducer was added to the other group, which were repressed by growth on minimal media with glucose as sole carbon source. Both groups were assayed for general protein production by BioRad Assay, following cultivation, filtering to separate the mycelia, freeze drying, grinding and protein extraction with 20 mM Tris/HCl at pH 7.

To test for production of cellulase, plates of Agar medium containing cellulase (9 g/Lt, carboxymethylcellulose) were prepared, and small pieces of glass fibre filter material, isolated from one another, and 75 ug of total protein from one of the transconjugants was added to each of the filters. The plates were incubated overnight at 37°C. The filters were then removed, and the plates stained with congo red to determine the locations where cellulase had been present in the total protein on the filters, as evidenced by the breakdown of cellulase in the agar medium below. The plates were de-stained, by washing with 5M NaCl in water, to detect the differences visibly.

Of four transformants induced with threonine and fructose, three clearly showed the presence of cellulase in the total protein product. The non-induced, glucose repressed transformants did not show evidence of cellulase production.

Three control transformants were also prepared from the same vector system and strains, but omitting the promoter sequence. None of them produced cellulase, with or without inducers. The presence of C. fimi endoglucanase coding region was verified by the fact that medium from threonine-induced transformed strains showed reactivity with a monoclonal antibody raised against C. fimi endoglucanase. This monoclonal antibody showed no cross-reactivity with endogenous A. nidulans proteins in control strains.

- 22 -

Example 2 - Transformation of *A. nidulans* using pALCALAMY
ATCC 53380

The vector construct pALCALAMY was prepared as indicated in Figure 16, using standard routine ligation and restriction techniques. In particular, and with reference to Figure 16 vector pALCAL containing a Hind III-EcoRI segment in which the *A. nidulans* alcohol dehydrogenase 1 promoter 22 is located (as described previously), was cut at its EcoRI site in order to insert the coding region of the wheat α -amylase gene 72 contained within an EcoRI-EcoRI fragment defined on plasmid p501 (see S.J. Rothstein et al, Nature, 308, 662-665 (1984)). As wheat α -amylase is a naturally secreted protein, its coding region 72 contains a signal peptide coding region 76 and a segment 78 which encodes mature, secreted α -amylase. Ligation of coding region 72 contained in the EcoRI-EcoRI segment of p501 within the EcoRI-cut site of pALCAL provides plasmid pALCALAMY in which the AlcA promoter 22 is operatively associated with the α -amylase coding region. The correct orientation of the p501-derived α -amylase coding region within pALCALAMY is confirmed by sequencing across the ligation site according to standard procedures. The nucleotide sequence of the promoter/coding region junction is shown in Figure 17.

A. nidulans may be transformed by the procedure described in example 1, samples of extracellular medium being taken from and applied to glass fibre filter papers placed on 1% soluble starch agar. The filters are then removed after 8 hours at 37°C and inverted onto beakers containing solid iodine (in a 50°C water bath). Clear patches indicate starch degradation while the remaining starch turns a deep purple, thereby confirming the presence of secreted α -amylase.

In examples 3-12 which follow, vectors are provided in which a secretion signal peptide coding region is introduced in the vector in order to obtain secretion of a foreign protein from a filamentous fungus transformed by the entire vector.

Example 3 - Production of Plasmid pGL2, an intermediate vector

A) Source of promoter and signal peptide sequence

The glucoamylase gene of A. niger was isolated by probing a gene bank derived from DNA available in a strain of this microorganism on deposit with ATCC under catalogue number 22343. The probing was conducted using oligonucleotide probes prepared with Biosearch oligonucleotide synthesis equipment and with knowledge of the published amino acid sequence of the glucoamylase protein. The amino acid sequence data was "reverse translated" to nucleotide sequence data and the probes synthesized. The particular gene bank probed was a Sau 3A partial digest of the A. niger DNA described above cloned into the Bam HI site of the commercially available plasmid pUC12 which is both viable in and replicable in E. Coli.

A Hind III -Bgl II piece of DNA containing the glucoamylase gene was subcloned into pUC12. Subsequently, the location of the desired promoter region, signal peptide coding region and protein coding region of the glucoamylase was identified within pUC12 containing the sub-cloned fragment. The EcoRI/EcoRI fragment (see Figure 4) was shown to contain a long, open translation reading frame when it was sequenced and the sequence data was analyzed using the University of Wisconsin sequence analysis programmes.

Results of analysis of the nucleotide sequence of part of the region of the glucoamylase gene between the 5' Eco RI site and BssH II 3' site within the Hind III - Bgl II fragment are shown in Figure 6. This region contains the glucoamylase promoter and the signal peptide coding region.

Within this fragment i.e. at nucleotides 97-102 is a "TATA box" 48 which provides a site required by many eukaryotic promoter regions for accurate initiation of transcription (probably an RNA polymerase II binding site). Accordingly, the

- 24 -

presence of at least a portion of the promoter region is confirmed. Further, it is predictable from analogy with other known promoter regions that all the functional essentials are likely to be contained within a sequence of about 1,000 bases in length and most likely within the first 200 - bases upstream of the start codon for the coding region i.e. nucleotides 206 - 208 or "ATG" 49, the codon for methionine. Thus, the promoter and transcript leader terminate at nucleotide 205. The identity of the beginning of the promoter region is less crucial although the promoter region must contain the RNA polymerase II binding site and all other features required for its function. Thus, whereas the Eco RI-Eco RI sequence is believed to represent the entire promoter region of the glucoamylase gene, the fragment used in plasmid pGL2 contains this fragment in the much larger Hind III - BamH I/Bgl II segment to ensure that the entire promoter region is properly included in the resultant plasmid.

On the basis that the amino acid sequence of mature glucoamylase is known (see Svensson et al, "Characterization of two forms of glucoamylase from *Aspergillus niger*", Carlsberg Res. Commun, 47, 55-69 (1982)), a nucleotide sequence of the signal peptide can be determined accurately. The signal peptide coding region of genes encoding secreted proteins is known to initiate with the methionine residue encoded by the ATG codon 49. Determination of a sufficient initial portion of the nucleotide sequence beyond i.e. 3' of the ATG codon provides information from which the amino acid sequence of that portion may be determined. By comparison of this amino acid sequence with the published amino acid sequence, the signal peptide can be identified as that portion of the glucoamylase gene which has no counterpart in the published sequence with which it was compared. The glucoamylase signal peptide coding region defined herein was previously confirmed using this method.

By the above methods, the Hind III - Bam HI/Bgl II fragment resulting from Sau 3A partial digestion and incorporated into pUC12 was confirmed to contain the following

features of the glucoamylase gene: an initial, perhaps non-relevant section, the promoter region, the signal peptide coding region and the remaining portion of the coding region. This fragment, inserted into the pUC12 plasmid by scission with Hind III and Bam HI/Bgl II and ligation appears schematically in Figure 4 as plasmid pGL1. This plasmid contains all of the features necessary for replication and the like in order to remain selectable and replicable in E. Coli.

B) Construction of Plasmid pGL2

Using pGL1 as a precursor, plasmid vector pGL2 can be formed as shown in Figure 4. The restriction site BssH II near the 3' end of the signal sequence 42, is utilized together with the unique downstream Sst I site in order to insert a synthetic linker sequence A, B, or C defined in Figure 5 herein. Thus, pGL1 is cleaved with both BssH II and Sst I thereby removing the initial portion of the glucoamylase coding region 46 contained therein. Thereafter a selected one of the synthetic leader sequences A through C having been designed so as to be flanked by BssH II/Sst I compatible ends is inserted and ligated, thereby generating plasmid pGL2. Depending on which of the three linker sequences is used i.e. A, B or C, the resultant plasmid will hereinafter be identified as pGL2A, pGL2B or pGL2C, respectively.

The synthetic linker sequences identified herein are each equipped with unique Eco RV and Bgl II restriction sites, as shown in Figure 5, into which a desired protein coding region may be inserted. Once inserted, the resultant plasmid may be used to transform a host e.g. A. niger, A. nidulans and the like. The presence of the promoter region and the signal peptide coding region both of which are recognized by the host, provide a means whereby expression of the protein coding region and secretion of the protein so expressed is made possible.

Example 4 - Use of Plasmid pGL2 in creating pGL2BIFN

An example of the utility of the plasmid pGL2 is described below with reference to Figure 7, which shows schematically the construction of plasmid pGL2BIFN from pGL2B.

5 The plasmid pGL2B is prepared as described in general previously for pGL2 save that synthetic linker sequence "B" shown in Figure 5 is inserted specifically. The reference numeral 44 has accordingly been modified in Figure 7 to read "44B". In order to make available an opening in the vector
10 pGL2B, the plasmid is cut with Eco RV at the site internal to linker 44B. The scission results in blunt ends which may be ligated with a fragment flanked by blunt ends using ligases known to be useful for this specific purpose.

15 In the embodiment depicted in Figure 7, a fragment 60 containing the coding region of human interferon α -2 is inserted to create pGL2BIFN. Specifically, a Dde I - Bam HI fragment 60 containing the coding region coding for human interferon α -2 was excised from plasmid pN5H8 (not shown) on the basis of the known sequence and restriction map of this gene.

20 The plasmid pN5H8 combines known plasmid pAT153 with the interferon gene at a Bam HI site. The interferon gene therein is described by Slocomb, et. al., "High level expression of an interferon α -2 gene cloned in phage M13mp7 and subsequent purification with a monoclonal antibody" Proceedings
25 of the National Academy of Sciences, U.S.A., Vo. 79 pp 5455-5459 (1982)

30 In order to anneal the sticky ends of the interferon fragment into the cut Eco RV site of pGL2B, the sticky Dde I and Bam HI ends are filled using reverse transcriptase and ligated with an appropriate ligase according to techniques standard in the art.

The advantage of selecting linker sequence B for insertion into pGL2 is manifest from Figure 8 which shows the reading frame of the interferon α -2 coding region and its relationship with the recreated signal peptide sequence, in terms of nucleotide sequence and amino acid sequence, where appropriate.

Figure 8 shows a portion of the promoter region 40 5' of the signal sequence joined with a portion of the glucoamylase signal peptide sequence 42 beginning with the methionine codon ATG at 49 and ending with the lysine codon AAG at 50. In fact, although the signal peptide coding region extends one residue further i.e. to the CGC codon for arginine at 52, this latter residue is comprised by the synthetic linker sequence 44B engineered so as to compensate for the loss of the arginine residue during scission and ligation to insert the linker sequence. In this way, the genetic sequence of the signal remains undisturbed.

In a similar manner, the linker sequence provides for insertion of the interferon α -2 coding region without altering the reading frame thereof. With reference to Figures 7 and 8 cleavage of linker sequence 44B by Eco RV results in linker fragments 44B' and 44B" having blunt ends. Excision of the interferon α -2 coding region at Dde I site results, after filling in of the sticky ends created by the enzyme, in the desired nucleotide sequence without harming the sequence of that coding region. Ligation within the Eco RV-cleaved linker sequence of the interferon sequence filled at the Dde I site maintains the natural reading frame of the interferon coding region as evidenced by the triplet codon state between the linker portion 44B' and the interferon coding region 60. Had the linker A shown in Figure 5 been chosen, which bears one less nucleotide than the linker B, the entire reading frame would have been shifted by one nucleotide resulting in a nonsense sequence. By selection of synthetic linker B, codons are made available between the signal peptide sequence and the interferon

- 28 -

coding region which do not alter the reading frame of the coding region, when the blunt ended 1F α -2 fragment is oriented correctly. The correct orientation is selected by sequencing clones with inserts across the ligation junction.

5 Example 5 - Expression and Secretion from *A. nidulans*
 Transformed with pGL2BIFN ATCC 53371

10 The plasmid pGL2BIFN was cotransformed i.e. with a plasmid containing Arg B⁺ gene as described more fully in U.S. patent application serial No. 678,578 filed December 5, 1984 into an Arg B⁻ strain of *A. nidulans* with a separate plasmid containing an arg B selectable marker. Arg B⁺ transformants were selected of which 18 of 20 contained 1 - 100 copies of the human interferon α -2 coding region (as detected by Southern blot analysis).

15 Several transformants were grown on starch medium to induce the glucoamylase promoter and the extracellular medium was assayed for human 1F α -2 using the CellTech 1F α -2 assay kit.

20 All transformants exhibited some level of synthesis and secretion of assayable protein. Two controls, the host strain (not transformed) and one arg B⁺ transformant with no detectable human 1F α -2 DNA showed no detectable synthesis of 1F α -2 protein. In a separate experiment, transformation of A. niger, rather than A. nidulans, with pGL2B1FN using, mutatis
25 mutandis, the same procedure as described above, demonstrated the ability of A. niger to secrete 1F α -2.

Thus, although the promoter and signal regions of pGL2BIFN are derived from A. niger they are shown to be operative in both A. nidulans and A. niger.

30 In the present invention, use may be made of promoter regions other than the glucoamylase promoter region. Suitable

- 29 -

for use are the promoter regions of the alcohol dehydrogenase I gene and the aldehyde dehydrogenase gene, illustrated in Figures 1A and 1B.

Example 6 - Construction of Plasmid pALCA1S, ATCC 53368
an intermediate vector

For use with the present example, the alca promoter was employed as comprised within an 10.3 kb plasmid pDG6 deposited with ATCC within host E. Coli JM83 under accession number 53169. A plasmid map of pDG6 is shown in Figure 2 and, for ease of reference, in Figure 9 to which reference is now made, to illustrate another embodiment using the alca promoter.

pDG6 comprises, in its Hind III-EcoRI (first occurrence) segment, the promoter region 22 of the alca gene as well as a small 5' portion of the alca coding region 3' of the start codon, ligated to the endoglucanase coding region 30. pDG6 further comprises a multiple cloning site 62 downstream of the C. fimi endoglucanase coding region 20.

To retrieve the alca promoter region 22, pDG6 was cut with Pst I and Xho I removing the bulk of the endoglucanase coding region 30. In a second step, the linearized plasmid 64 was resected in one direction in a controlled manner with exonuclease III (which will resect from XhoI but not PstI-cut DNA ends) followed by tailoring with nuclease S1. The resection was timed so that the enzyme removed nucleotides to a position 50 bases 5' of the alca ATG codon, leaving the TATA box and messenger RNA start site intact.

Following resection, the vector 66 was religated (recircularized) creating vector 68 bearing Sal I-Xba I restriction sites immediately downstream of the promoter region 22. Cleavage of vector 68 with Sal I/Xba I permits introduction of a signal peptide coding region at an appropriate location within the vector.

The particular signal peptide coding region employed in the present example was synthesized to reproduce a characteristic signal peptide coding region identified according to standard procedures as described by G. Von Heijne in Eur. J. Biochem. 17-21, (1983). The synthetic signal was engineered so as to provide a 5' flanking sequence complementary to a Sal I cleavage site and a 3' flanking sequence enabling ligation with the Xba I restriction sequence.

The sequence of the synthetic secretion signal 68 is reproduced below:

Sal I

```

      TCGACATGTACCGGTTCTCTCGCCGTCATCTCGGCCTTCCTCGCCACTGCC'TTCGCCAAG
1  -----+-----+-----+-----+-----+-----+ 59
      GTACATGGCCAAGGAGCGGCAGTAGAGCCGGAAGGAGCGGTGACGGAAGCGGTTC

```

MetTyrArgPheLeuAlaValIleSerAlaPheLeuAlaThrAlaPheAlaLys

Xba I

```

      T
60 ----- 64
      AGATC

```

SerArg

The secretion signal per se begins with Met and ends with the fourth occurrence of Ala, as indicated by the arrow.

Once generated, the synthetic sequence 68 acting as signal is cloned into the Sal I-Xba I site of vector 70 resulting in plasmid pALCALS which contains alca promoter region 22, and synthetic peptide signal coding region 68. That the signal peptide coding region is inserted upstream of the multiple cloning site 62 is significant in that the site 62 allows for cloning of a variety of protein coding segments within this plasmid.

Accordingly, pALCALS constitutes a valuable embodiment of the present invention.

Example 7 - Construction of Plasmid pALCAlSIFN

As an example of the utility of pALCAlS, reference is made to Figure 10 showing creation of pALCAlSIFN. This plasmid comprises the promoter region 22 of the alca gene and the synthetic signal peptide coding region 68 both of which are derived from pALCAlS (Figure 9). In addition, it contains the coding region 60 coding for human interferon α -2 derived from pGL2BIFN.

To obtain the protein encoding segment, pGL2BIFN is cleaved with Eco RI and partially cleaved with Bgl II (because of the presence of internal Bgl II sites). Insertion of the protein coding region is accomplished by cleaving pALCAlS with Bam HI and Eco RI both of which are available in the multiple cloning site 62 and ligating this coding region therein, thereby creating pALCAlSIFN.

The nucleotide sequence of the resultant plasmid, from a site 1170 nucleotides downstream of Hind III to Eco RI is shown in Figure 11, indicating the relevant sites of restriction endonuclease digestion. It will be noted from sheet 3 of Figure 11 that the 1F α -2 coding region 60 is in proper reading frame with the synthetic signal peptide coding region 68.

Example 8 - Expression and Secretion from A. Nidulans Transformed with Plasmid ALCAlSIFN

The plasmid pALCAlSIFN prepared as described above was co-transformed with A. nidulans to provide an arg B selectable marker, the arg B⁺ transformants selected and checked for the presence of the human interferon α -2 coding region, then grown on a threonine-containing medium to induce the alca promoter, all as described in example 3 above. The extracellular medium was assayed for human IF-2 using Cell Tech IF α 2 assay kit. Eleven of twenty transformants showed secretion of interferon, induced in the presence of threonine, and repressed in the presence of glucose.

Example 9 - pGL2CENDO ATCC 53372

In accordance with the procedures described in the previous examples, there was constructed a vector plasmid designated pGL2CENDO, from plasmid pGL2C ATCC 53367, analagous to pGL2BIFN shown in Fig. 7, but containing the endoglucanase coding region in place of the interferon 2 coding region, and using the synthetic linker sequence "C" (Fig. 5) in place of linker sequence "B". A Bam HI fragment containing the C. fimi endogluconase coding region 30 was inserted into the Bgl II site of pGL2C. A. nidulans transformants were prepared with this vector plasmid, and showed starch regulated secretion of cellulase assayed as described in Example 1. The map of vector plasmid pGL2CENDO is shown in Fig. 12 of the accompanying drawings, in which 30 denotes the endoglucanase coding region (the endoglucanase coding region of Cellulomonas fimi, described in connection with Fig. 2 and Example 1), 42 denotes the signal peptide coding region of the glucoamylase gene and 40 denotes the promoter region of the glucoamylase gene. The nucleotide sequence is shown in Figure 13 and exemplifies that use of linker sequence C (Fig. 5) retains the reading frame of the signal peptide coding region 42 and the endoglucanase coding region 30.

Example 10 - Construction of Plasmid pALCALSENDO ATCC 53370

In accordance with the procedures described in the previous examples, there was constructed a vector plasmid designated pALCALSENDO by combining Eco RI - linearized plasmid pALCALS as described in example 5 (Fig. 9) with an Eco RI fragment derived from plasmid pDG5B (see Fig. 2) (pDG5 with the orientation of the Hind III fragment reversed in pUC12) and containing the endoglucanase coding region 30. The map of pALCALSENDO is shown in Figure 14 and the nucleotide sequence of its pertinent region is shown in Figure 15. In these figures, the promoter region derived from alca is designated by numeral 22, the synthetic signal peptide coding region is designated 68

and the endoglucanase coding region is designated by reference numeral 30.

Example 11 - Expression and Secretion from A. nidulans
Transformed with pALCAISENDO and pGL2CENDO

5 A. nidulans was co-transformed with an argB⁺ selectable marker and the plasmid pALCAISENDO or pGL2CENDO prepared as described above. Of the co-transformants obtained several showed varying levels of secretion of cellulase (i.e. endoglucanase) as assayed on carboxymethylcellulose plates and
10 the monoclonal antibody test systems as described in example 1. Both plasmid transformants showed secretion which was controlled by the linked promoter. Plasmid pGL2CENDO was induced by starch and pALCAISENDO was induced with threonine.

15 Example 12 - Expression and Secretion From A. niger
Transformed with pGL2CENDO

A. niger was cotransformed with an argB⁺ selectable marker and the plasmid pGL2CENDO. Several of the transformants showed varying levels of secretion of endoglucanase as assayed as described in example 1. This secretion was induced by the
20 presence of starch in the medium.

Example 13 - Increased Copy Number of Regulatory Genes

 In Aspergillus nidulans the alcA promoter is turned on in the presence of the appropriate inducer, such as ethanol, by the action of the gene product of alcR, the positive regulatory
25 gene for alcA.

 Evidence with multiple copy transformants (containing multiple alcA promoters) suggests that the alcR gene product limits the promoter function of the several alcA promoters requiring stimulation.

Increasing the copy number of the alcR gene increases the expression of alcR and relieves this situation. The evidence for this is as follows:

5 Transformants with multiple copies of the alca promoter fused to its own coding region (ADH I) in a multiple alcR background (which has been shown to overproduce alcR messenger RNA) do not grow well on ethanol. This is probably due to rapid accumulation of aldehydes, the product of ADH breakdown of ethanol. ADH activity in these strains is high. The increased
10 activity of ADH due to increased copy number probably accounts for these observations.

Transformants with multiple copies of the alca promoter fused to interferon α -2 in a multiple alcR background produce significantly higher levels of secreted interferon. In these
15 strains, unlike those with single copy alcR, many more of the alca promoters have access to the alcR regulatory protein.

Thus, preferred embodiments of the present invention provide means for introducing a coding region into a filamentous fungus host which, when transformed, will secrete the desired
20 protein. Particularly useful intermediate plasmids for this purpose are pALCA1S and pGL2 (A, B or C).

Useful transformation vectors created from these plasmids include pALCA1SIFN, pGL2BIFN, pALCA1SENO and pGL2CENDO. Cultures of each of these and other plasmids
25 mentioned herein are currently maintained in a permanently viable state at the laboratories of Allelix Inc., 6850 Goreway Drive, Mississauga, Ontario, Canada. The plasmids will be maintained in this condition throughout the pendency of this patent application and, during that time, will be made available
30 to authorized persons. After issue of a patent on this application, these plasmids will be available from the ATCC depository recognized under the Budapest Treaty, without restriction. The accession numbers of the respective deposits appear in the table below:

- 35 -

	<u>Plasmid</u>	<u>Host</u>	<u>Accession #</u>	<u>Deposit Date</u>
	pDG6	E. Coli JM83	53169	June 7, 1985
	pGL2A	"	53365	Dec. 16, 1985
	pGL2B	"	53366	"
5	pGL2C	"	53367	"
	pALCA1S	"	53368	"
	pALCA1SEND0	"	53370	"
"	pALCA1SIFN	"	53369	"
	pGL2B1FN	"	53371	"
10	pGL2CENDO	"	53372	"
"	pALCA1AMY	"	53380	Dec. 20, 1985

CLAIMS:

1. A DNA construct comprising a promoter operative in a filamentous fungus to promote transcription of a coding region when operatively associated therewith.
2. The construct according to claim 1 wherein said promoter is native to a filamentous fungus gene.
3. The construct according to claim 2 wherein said promoter is native to an Aspergillus sp gene.
4. The construct according to claim 3 wherein said promoter is native to an Aspergillus nidulans gene.
5. The construct according to claim 4 wherein said promoter is the promoter of the alcohol dehydrogenase I gene of Aspergillus nidulans.
6. The construct according to claim 4 wherein said promoter is the promoter of the aldehyde dehydrogenase gene of Aspergillus nidulans.
7. The construct according to claim 3 wherein said promoter is native to an Aspergillus niger gene.
8. The construct according to claim 7 wherein said promoter is the promoter of the glucoamylase gene of Aspergillus niger.
9. A DNA sequence operative in a filamentous fungus to signal secretion of a protein encoded by a protein coding region when said protein coding region is in proper translational reading frame with said DNA sequence, said DNA sequence having the nucleotide sequence substantially as defined below:

10. A DNA construct comprising a promoter operative in a filamentous fungus in operative association with a coding region encoding a protein.
11. The construct of claim 10 wherein the coding region is foreign to the promoter.
12. The construct of claim 11 wherein the promoter is native to a filamentous fungus.
13. The construct of claim 11 wherein the promoter is native to a filamentous fungus of Aspergillus sp.
14. The construct of claim 11 wherein the promoter is native to Aspergillus nidulans or Aspergillus niger.
15. The construct of claim 11 wherein the promoter is selected from the promoter of the alcohol dehydrogenase I gene of A. nidulans, the promoter of the aldehyde dehydrogenase gene of A. nidulans and the promoter of the glucoamylase gene of A. niger.
16. The construct according to claim 10 which comprises the promoter of the alcohol dehydrogenase I gene of A. nidulans in operative association with the endoglucanase coding region of C. fimi.
17. The construct according to claim 16 as contained on plasmid pDG6 having ATCC catalogue number 53169.
18. The DNA expression vector according to claim 11 wherein the coding region naturally contains an integral signal peptide coding region.

19. The construct of claim 18 wherein the promoter is native to a filamentous fungus.
20. The construct of claim 18 wherein the promoter is native to a filamentous fungus of Aspergillus sp.
21. The construct of claim 18 wherein the promoter is native to Aspergillus nidulans or Aspergillus niger.
22. The construct of claim 18 wherein the promoter is selected from the promoter of the alcohol dehydrogenase I gene of A. nidulans, the promoter of the aldehyde dehydrogenase gene of A. nidulans and the promoter of the glucoamylase gene of A. niger.
23. The construct according to claim 18 which comprises the promoter of the alcohol dehydrogenase I gene of Aspergillus nidulans in operative association with the wheat α -amylase coding contained on plasmid p501.
24. The construct according to claim 23 as contained on plasmid pALCALAMY having ATCC catalogue number 53380.
25. A DNA construct comprising, in a 5' to 3' direction:
- a) a promoter operative in a filamentous fungus;
 - b) a signal peptide coding region in operative association with the promoter, said signal peptide coding region encoding a polypeptide which functions in a filamentous fungus to signal secretion of a protein fused functionally thereto; and
 - c) a segment of DNA linked with said signal peptide coding region and defining a restriction site within which a protein coding region may be inserted such that said protein coding region is in proper translational reading frame with said signal peptide coding region.

28. The construct according to claim 27 wherein said signal peptide coding region is native to a gene coding for a protein secreted from Aspergillus niger or Aspergillus nidulans.

30. The construct according to claim 29 wherein the promoter is the promoter of the glucoamylase gene of Aspergillus niger.

32. The construct according to claim 26 wherein said signal peptide coding region has a nucleotide sequence characteristic of signal peptide coding regions operative in a filamentous fungus.

5' GTCGACATGTACCGGTTCTCGCCGTCATCTCGGCCTTCT
---+-----+-----+-----+

CGCCACTGCC 3'
-----+-----

34. The construct according to claim 33 wherein the promoter is the promoter of the alcohol dehydrogenase I gene of Aspergillus nidulans.

35. The construct according to claim 34 as contained on plasmid pALCA1S having ATCC catalogue number 53368.

36. A DNA construct comprising, in a 5' to 3' direction:

a) a promoter operative in a filamentous fungus;

b) a signal peptide coding region in operative association with said promoter, said signal peptide coding region encoding a signal peptide which serves to signal secretion from a filamentous fungus of a protein when fused functionally to the protein; and

c) a protein coding region which codes for a protein said protein coding region being linked in proper translational reading frame with said signal peptide coding region.

37. The construct according to claim 36 wherein at least one of said promoter, said signal peptide coding region and said protein coding region is foreign to the other or others of said promoter, said signal peptide coding region and said protein coding region.

38. The construct according to claim 37 wherein said protein coding region is foreign to said signal peptide coding region and to said promoter.

39. The construct according to claim 38 wherein said promoter and said signal peptide coding region are native to a filamentous fungus.

40. The construct according to claim 38 wherein said promoter and signal peptide coding region are native to Aspergillus sp.

- 41 -

41. The construct according to claim 40 wherein the promoter is selected from the promoter of the alcohol dehydrogenase I gene of Aspergillus nidulans, the promoter of the aldehyde dehydrogenase gene of Aspergillus nidulans and the promoter of the glucoamylase gene of Aspergillus niger.

42. The construct according to claim 40 wherein the promoter is the promoter of the glucoamylase gene of Aspergillus niger and the signal peptide coding region is the signal peptide coding region of said gene.

43. The construct according to claim 42 wherein the protein coding region is the endoglucanase coding region of the C. fimi endoglucanase gene.

44. The construct according to claim 43 as contained on plasmid pGL2CENDO having ATCC catalogue number 53372..

45. The construct according to claim 42 wherein the protein coding region is the interferon α -2 coding region of the human interferon α -2 gene.

46. The construct according to claim 45 as contained on plasmid pGL2BIFN having ATCC catalogue number 53371.

47. The construct according to claim 38 wherein said signal peptide coding region has a nucleotide sequence characteristic of signal peptide coding regions operative in and native to filamentous fungi.

48. The construct according to claim 47 wherein the promoter is native to a filamentous fungus.

49. The construct according to claim 48 wherein the signal peptide coding region has a nucleotide sequence substantially as shown below:

57. The DNA expression vector according to claim 56 wherein the host which may be transformed therewith is of Aspergillus sp.

58. The DNA expression vector according to claim 56 wherein the host which may be transformed therewith is Aspergillus nidulans.

59. The DNA expression vector according to claim 56 wherein the host which may be transformed therewith is Aspergillus niger.

60. The DNA expression vector according to claim 57 in plasmid form.

61. The DNA expression vector according to claim 57 which is plasmid pDG6 ATCC 53169.

62. The DNA expression vector according to claim 57 which is plasmid pALCALAMY ATCC 53380.

63. The DNA expression vector according to claim 57 which is plasmid pALCALSENDO ATCC 53370.

64. The DNA expression vector according to claim 57 which is plasmid pALCALSIFN ATCC 53369.

65. The DNA expression vector according to claim 57 which is plasmid pGL2BIFN ATCC 53371.

66. The DNA expression vector according to claim 57 which is plasmid pGL2CENDO ATCC 53372.

67. A filamentous fungal cell having the capacity to express proteins foreign thereto.

68. A filamentous fungus cell transformed by a DNA construct as defined in any one of claims 10-24 and 36-55.

69. Transformants according to claim 68 which are of the species Aspergillus sp.
70. Transformants according to claim 68 which are of the species Aspergillus nidulans.
71. Transformants according to claim 68 which are of the species Aspergillus niger.
72. Transformants according to claim 68 containing multiple copies of said constructs.
73. Transformants according to claim 72 wherein the transformants express greater than normal levels of regulatory gene products.
74. Transformants according to claim 73 wherein the greater than normal levels of regulatory gene products results from multiple copies of genes which regulate the transcription promoting function of a promoter comprised by said construct.
75. Transformants according to claim 74 containing multiple copies of a construct comprising the promoter of the alcohol dehydrogenase I gene and multiple copies of the alcR gene.
76. A process for expressing a protein in a filamentous fungus cell as defined in claims 68-75 which comprises culturing the fungus cells under appropriate growth conditions.
77. A process as defined in claim 76 wherein the fungus is cultured under conditions in which the presence of substances which induce or repress the transcription promoting function of the promoter is controlled.
78. A filamentous fungal cell having the capacity to secrete proteins foreign thereto.

79. A filamentous fungal cell transformed by a DNA construct as defined in any one of claims 18-24 and 36-55.
80. Transformants according to claim 79 which are of the species Aspergillus sp.
81. Transformants according to claim 79 which are of the species Aspergillus nidulans.
82. Transformants according to claim 79 which are of the species Aspergillus niger.
83. Transformants according to claim 79 containing multiple copies of said constructs.
84. Transformants according to claim 83 wherein the transformants express greater than normal levels of regulatory gene products.
85. Transformants according to claim 84 wherein the greater than normal levels of regulatory gene products results from multiple copies of genes which regulate the transcription promoting function of a promoter comprised by said construct.
86. Transformants according to claim 85 containing multiple copies of a construct comprising the promoter of the alcohol dehydrogenase I gene and multiple copies of the alcR gene.
87. A process for obtaining expression of and secretion of a protein in a filamentous fungus cell as defined in claims 79-86 which comprises culturing the fungus under appropriate growth conditions.
88. The process according to claim 87 wherein the fungus cell is cultured under conditions in which the presence of substances which induce or repress the transcription promoting function of the promoter is controlled.

1/25

1
GGATACAGTTGGGCATTTCTAGGGCTGAATGGGAAGGAGAGAGTTTTGAAATAGGCGTTC
-----+-----+-----+-----+-----+-----+-----+

CGTTCTGCTTAGGGTATTTGGAACAATCAATGTTCAATGTACATTTAATCCACGATTTT
-----+-----+-----+-----+-----+-----+-----+

ATAAAACGTCATCCTTTGCCCTCCCTTCTTATTTGCCAATACCAAAAATCTTACTCCAGT
-----+-----+-----+-----+-----+-----+-----+

GGTTCGGTAATCGCAGAGTTAAATCTGGGCTCGGTGGCAGATCTGCGATCGTCCATAACC
-----+-----+-----+-----+-----+-----+-----+
17

G TTCAGATGTTGATTGGAAC TGGGTGGGGTAGACAGCTCCGAAGACCGAGTGAACGTATA
-----+-----+-----+-----+-----+-----+-----+

CCTAAGACACTTTGACACGGCCGGAACACTGTAAGTCCCTTCGTATTTCTCCGCCTGTGT
-----+-----+-----+-----+-----+-----+-----+

GGAGCTACCATCCAATAACCCCCAGCTGAAAAAGCTGATTGTCGATAGTTGTGATAGTTC
-----+-----+-----+-----+-----+-----+-----+

CCACTTGTCGGTCCGCATCGGCATCCGCAGCTCGGGATAGTTCCGACCTAGGATTGGATG
-----+-----+-----+-----+-----+-----+-----+

CATGCGGAACCGCACGAGGGCGGGCGGAAATTGACACACCACTCCTCTCCACGCAGCCG
-----+-----+-----+-----+-----+-----+-----+
14

TTCAAGAGGTACGCGTATAGAGCCGTATAGAGCAGAGACGGAGCACTTTCTGGTACTGTC
-----+-----+-----+-----+-----+-----+-----+

CGCACGGGATGTCCGCACGGAGAGCCACAAACGAGCGGGGCCCCGTACGTGCTCTCCTAC
-----+-----+-----+-----+-----+-----+-----+

CCCAGGATCGCATCCTCGCATAGCTGAACATCTATATAAAGACCCCCAAGGTTCAGTC
-----+-----+-----+-----+-----+-----+-----+
16

TCACCAACATCATCAACCAACAATCAACAGTTCTCTACTCAGTTAATTAGAACACTTCCA
-----+-----+-----+-----+-----+-----+-----+

ATCCTATCACCTCGCCTCAAAATGTGCATCCCCACTATGCAATGGGCCAGGTGCGCGAG
-----+-----+-----+-----+-----+-----+-----+
12 13 10
Met Cys Ile Pro Thr Met Gln Trp Ala Gln Val Ala Glu

2/25

841

AAGGTCGGCGGCCCCGCTCGTCTACAAGCAGATCCCCGTCCCTAAGCCCCGTCCCGACCAG
-----+-----+-----+-----+-----+-----+-----+
LysValGlyGlyProLeuValTyrLysGlnIleProValProLysProGlyProAspGln

ATCCTTGTGAAGATCCGCTACTCTGGGGTTTGCCACACCGACCTACACGCTATGATGGGT
-----+-----+-----+-----+-----+-----+-----+
IleLeuValLysIleArgTyrSerGlyValCysHisThrAspLeuHisAlaMetMetGly

CACTGGCCAATCCCCGTCAAAATGCCGCTCGTCCGTGGGCACGAAGGAGCAGGAATCGTC
-----+-----+-----+-----+-----+-----+-----+
HisTrpProIleProValLysMetProLeuValGlyGlyHisGluGlyAlaGlyIleVal

GTGGCAAAGGGCGAACTGGTCCACGAATTTCGAGATCGGCGACCAAGCTGGCATCAAATGG
-----+-----+-----+-----+-----+-----+-----+
ValAlaLysGlyGluLeuValHisGluPheGluIleGlyAspGlnAlaGlyIleLysTrp

CTTAATGGTTCCTGCGGAGAGTGGGAGTTCTGCCGCCAATCGGACGACCCCTCTGTGCA
-----+-----+-----+-----+-----+-----+-----+
LeuAsnGlySerCysGlyGluCysGluPheCysArgGlnSerAspAspProLeuCysAla

CGCGCCCAGCTCTCTGGGTATACTGTTGACGGCACGTTCCAGCAGTATGCGCTCGGAAAG
-----+-----+-----+-----+-----+-----+-----+
ArgAlaGlnLeuSerGlyTyrThrValAspGlyThrPheGlnGlnTyrAlaLeuGlyLys

GCGAGTCATCGCTCGAAGATCCCTGCGGGCGTTCCGGTGGATGCCGCGGCCCCAGTACTC
-----+-----+-----+-----+-----+-----+-----+
AlaSerHisAlaSerLysIleProAlaGlyValProValAspAlaAlaAlaProValLeu

TGTGCCGGTATTACAGTGTACAAGGGATTGAAAGAGGCCGGGGTCCGGCCGGGCCAGACC
-----+-----+-----+-----+-----+-----+-----+
CysAlaGlyIleThrValTyrLysGlyLeuLysGluAlaGlyValArgProGlyGlnThr

GTGGCGATCGTGGGTGCCGGTGGCGGCCTGGGATCCCTTGACACAGCAGTATGCGAAGGCG
-----+-----+-----+-----+-----+-----+-----+
ValAlaIleValGlyAlaGlyGlyGlyLeuGlySerLeuAlaGlnGlnTyrAlaLysAla

ATGGGGATCAGGGTTGTGCGCGTTCGATGGGGGAGATGAGAAGCGGGCCATGTGTGAGTCG
-----+-----+-----+-----+-----+-----+-----+
MetGlyIleArgValValAlaValAspGlyGlyAspGluLysArgAlaMetCysGluSer

CTTGGAACAGAGGTATGTACATGTTCTCAATCTCAGGAAGCAAGCAACTGACCTGGACAG
-----+-----+-----+-----+-----+-----+-----+
LeuGlyThrGlu.....IVS 1.....

ACATATGTCGACTTCACCAAATCTAAAGACCTGGTGGCCGACGTCAGGCACGGACGCGGA
-----+-----+-----+-----+-----+-----+-----+
ThrTyrValAspPheThrLysSerLysAspLeuValAlaAspValArgHisGlyArgGly

3/25

1561

TGTCTGGGTGCGCACGCGGTGATCCTGCTTGCCGTGTCAGAGAAGCCCTTCCAGCAAGCC
-----+-----+-----+-----+-----+-----+-----+
Cys Leu Gly Ala His Ala Val Ile Leu Leu Ala Val Ser Glu Lys Pro Phe Gln Gln Ala

ACTGAATATGTGCGCTCTCGCGGGACAATTGTTGCTATTGGCTTGCCGCCAGATGCGTAC
-----+-----+-----+-----+-----+-----+-----+
Thr Glu Tyr Val Arg Ser Arg Gly Thr Ile Val Ala Ile Gly Leu Pro Pro Asp Ala Tyr

CTCAAGGCCCTGTGATCAACACAGTTGTTTCGCATGATCACTATCAAGGGCAGCTACGTT
-----+-----+-----+-----+-----+-----+-----+
Leu Lys Ala Pro Val Ile Asn Thr Val Val Arg Met Ile Thr Ile Lys Gly Ser Tyr Val

GGAAACCGACAGGACGGTGTTCGAGGCTCTGGACTTCTTCGCTCGCGGCCTGATCAAGGCT
-----+-----+-----+-----+-----+-----+-----+
Gly Asn Arg Gln Asp Gly Val Glu Ala Leu Asp Phe Phe Ala Arg Gly Leu Ile Lys Ala

CCGTTCAAGACGGCTCCTCTGAAGGATCTGCCGAAGATTTACGAGCTTATGGGTGCGTTG
-----+-----+-----+-----+-----+-----+-----+
Pro Phe Lys Thr Ala Pro Leu Lys Asp Leu Pro Lys Ile Tyr Glu Leu Met.....

ACTCCCATATCCGATGTTCAATTCTCTTTGCGCGATATATTTAGATACTAATGGCTTGCA
-----+-----+-----+-----+-----+-----+-----+
.....IVS 2.....

GAACAAGGCAGAATCGCCGGTCGTTATGTGCTAGAGATGCCAGAATAAGCGTTTCAACGC
-----+-----+-----+-----+-----+-----+-----+
Glu Gln Gly Arg Ile Ala Gly Arg Tyr Val Leu Glu Met Pro Glu End

10

CCACGGGCTGGA ACTACAAACACAATCGTCAGATGTTTCATGTTTATGATGTCCATGCTT
-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

GATATCTTTGTATATAGTTTTCAATCAAGTGGTACAATGATTTTGGCCTTG GTTCAACCG
-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

ATTTCCCTTCCACTTTTCTAGCCTGATACGGATAGCACTTGTAAGCAATCATAAACAC
-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

AGACTGGATAGACTGGGAAGTATAGGTATTATGGTAGCCAAATCCGATGATTTGACTTTG
-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

AATCAAGGCTCAA ACTAGCCCTCACCGTCTCTTTTTCGGTATTTTTAGTCGATTCACG
-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

2280

1

840

5/25

841

TCGCCCTCCTGATTTTTTTTGTGTTGTGTTTATTAACGATCATTGGGTGTAGGTTTCATCA
-----+-----+-----+-----+-----+-----+-----+
.....IVS 1.....Phe Ile A

ACAACGAGTTCGTGAAGGGCGTTGAGGGCAAGACCTTCCAGGTCATCAACCCCTCCAACG
-----+-----+-----+-----+-----+-----+-----+
snAsnGluPheValLysGlyValGluGlyLysThrPheGlnValIleAsnProSerAsnG

AGAAGGTCATCACCTCCGTCCACGAAGCCACCGAGAAGGATGTTGATGTCGCCGTCGCTG
-----+-----+-----+-----+-----+-----+-----+
luLysValIleThrSerValHisGluAlaThrGluLysAspValAspValAlaValAlaA

CTGCCCGTGC TGCTTTGAgGGGCCATGGCGCCAGGTCACCCCTCTGAGCGTGGCATT
-----+-----+-----+-----+-----+-----+-----+
laAlaArgAlaAlaPheGluGlyProTrpArgGlnValThrProSerGluArgGlyIleL

TGATCAACAAGCTGGCGGATCTGATGGAGCGTGATATCGACACCCTCGCCGCTATCGAGT
-----+-----+-----+-----+-----+-----+-----+
euIleAsnLysLeuAlaAspLeuMetGluArgAspIleAspThrLeuAlaAlaIleGluS

CTCTCGACAACGGCAAGGCTTTCACCATGGCCAAGGTCGATCTTGCCAACTCcATTGGTT
-----+-----+-----+-----+-----+-----+-----+
erLeuAspAsnGlyLysAlaPheThrMetAlaLysValAspLeuAlaAsnSerIleGlyC

GCTTGCGATACTACGCTGGCTGGGCGGACAAGATTCACGGTCAGACCATTGACACCAACC
-----+-----+-----+-----+-----+-----+-----+
ysLeuArgTyrTyrAlaGlyTrpAlaAspLysIleHisGlyGlnThrIleAspThrAsnP

CCGAGACTCTTACCTACACCCGCCACGAGCCCGTTGGTGTGTTGCGGTCAGATCATCCCT
-----+-----+-----+-----+-----+-----+-----+
roGluThrLeuThrTyrThrArgHisGluProValGlyValCysGlyGlnIleIleProT

GGAACCTCCCCCTTCTGATGTGGTCCTGGAAGATTGGACCCGCTGTTGCCGCTGGTAACA
-----+-----+-----+-----+-----+-----+-----+
rpAsnPheProLeuLeuMetTrpSerTrpLysIleGlyProAlaValAlaAlaGlyAsnT

CTGTTGTCTCAAGACCGCCCAGCAGACCCCTCTCTCCGCCCTTTACGCTGCTAAGCTGA
-----+-----+-----+-----+-----+-----+-----+
hrValValLeuLysThrAlaGlnGlnThrProLeuSerAlaLeuTyrAlaAlaLysLeuI

TCAAGGAGGCTCCAttCCCCGCTGGTGTGATCAACGTCATCTCTGGCTTTGGCCGTACCG
-----+-----+-----+-----+-----+-----+-----+
leLysGluAlaProPheProAlaGlyValIleAsnValIleSerGlyPheGlyArgThrA

CTGGTGCTGCCATCTCCAGCCACATGGACATTGACAAGGTTGCCTTCACTGGCTCTACTC
-----+-----+-----+-----+-----+-----+-----+
laGlyAlaAlaIleSerSerHisMetAspIleAspLysValAlaPheThrGlySerThrL

6/25

1561

TTGTTGGACCTACCA_tCCTGCAGGCCGCTGCTAAGAGCAACCTGAAGAAGGTCACCTCTTG
-----+-----+-----+-----+-----+-----+-----+
euValGlyProThrIleLeuGlnAlaAlaAlaLysSerAsnLeuLysLysValThrLeuG

AGCTCGGTGGCAAGTCTCCCAACATCGTCTTTGATGATGCTGACATTGACAACGCCATTT
-----+-----+-----+-----+-----+-----+-----+
luLeuGlyGlyLysSerProAsnIleValPheAspAspAlaAspIleAspAsnAlaIleS

CCTGGGCCAACTTTGGTATCTTCTTCAACCACGGCCAGTGCTGCTGTGCTGGATCCCGTA
-----+-----+-----+-----+-----+-----+-----+
erTrpAlaAsnPheGlyIlePhePheAsnHisGlyGlnCysCysCysAlaGlySerArgI

TCCTGGTCCAGGAGGGCATCTACGACAAGTTCGTGCGCCGCTTCAAGGAGCGTGCCCAGA
-----+-----+-----+-----+-----+-----+-----+
leLeuValGlnGluGlyIleTyrAspLysPheValAlaArgPheLysGluArgAlaGlnL

AGAACAAGGTCCGAAACCCCTTCGAGCAGGACACCTTCCAGGGTCCCCAGGTTTCCCAGC
-----+-----+-----+-----+-----+-----+-----+
ysAsnLysValGlyAsnProPheGluGlnAspThrPheGlnGlyProGlnValSerGlnL

TCCAGTTCGACCGTATCATGGAGTACATCAACCACGGCAAGAAGGCTGGTGCTACCGTCG
-----+-----+-----+-----+-----+-----+-----+
euGlnPheAspArgIleMetGluTyrIleAsnHisGlyLysLysAlaGlyAlaThrValA

CCACCGGTGGTGACCGCCACGGCAACGAGGGTTACTTCATCCAGCCTACTGTCTTCACAG
-----+-----+-----+-----+-----+-----+-----+
laThrGlyGlyAspArgHisGlyAsnGluGlyTyrPheIleGlnProThrValPheThrA

ACGTCACTTCCGACATGAAGATTGCCCAGGAGGAGATCTTCGGTCCCTGTCGTCACCTATCC
-----+-----+-----+-----+-----+-----+-----+
spValThrSerAspMetLysIleAlaGlnGluGluIlePheGlyProValValThrIleG

AGAAGTTCAAGGATGTGGCTGAGGCTATCAAGATCGGCAACTCGACCGACTACGGTACGT
-----+-----+-----+-----+-----+-----+-----+
lnLysPheLysAspValAlaGluAlaIleLysIleGlyAsnSerThrAspTyr.....

CTATCTTTTCTGGTCTTTGCCGATATTTTGTGCTAACATACGCACAGGTCTTGCTGCTG
-----+-----+-----+-----+-----+-----+-----+
.....IVS 2.....GlyLeuAlaAlaA

CCGTGCACACAAAGAACGTCAACACCGCCATTTCGGTGTCCAACGCTCTGAAGGCTGGTA
-----+-----+-----+-----+-----+-----+-----+
laValHisThrLysAsnValAsnThrAlaIleArgValSerAsnAlaLeuLysAlaGlyT

CCGTCTGGATCAACAACACTACAACATGATCTCGTACCAGGCTCCCTTCGGTGGCTTCAAGC
-----+-----+-----+-----+-----+-----+-----+
hrValTrpIleAsnAsnTyrAsnMetIleSerTyrGlnAlaProPheGlyGlyPheLysG

FIG.1 B sheet 3

2280

7/25

2281

AGTCCGGTCTCGGCCGTGAGCTTGGCTCTTACGCTCTTGAGAACTACACACAGATCAAGA

-----+-----+-----+-----+-----+-----+-----+
InSerGlyLeuGlyArgGluLeuGlySerTyrAlaLeuGluAsnTyrThrGlnIleLysT

10'

CGGTGCACTACCGCCTGGGTGATGCTCTTTTCGCTTAAAGCTAATTGTATGATTTGATGA

-----+-----+-----+-----+-----+-----+-----+
hrValHisTyrArgLeuGlyAspAlaLeuPheAlaEnd

2400

AATTGCGAATACAAGTTGGATATATCCTGTGTGCTACGGCACTGGTTCAAATTGCTTCTT

-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

GTGCAGCAACCATGTGACTCATGTAAAACATATCAGATAACCCCGGATACGATTTTACGA

-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

TTTTTTAGATTTGCTTTTATCGTAGCGTCCACTTATCCTCGTCCGGCCAAGCACAAAACC

-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

TATGGCTATCTTCAGCACGCCGCGATCCTGAACCGTAGCTGGATTGGAAATCCGAAATCA

-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

ACTGCCCCGCAGCCACCGACACTCGGGCTCCGGGCAAGTCCCCGCGAAATCCCTCACCAC

-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

2700

FIG.1B sheet 4

8/25

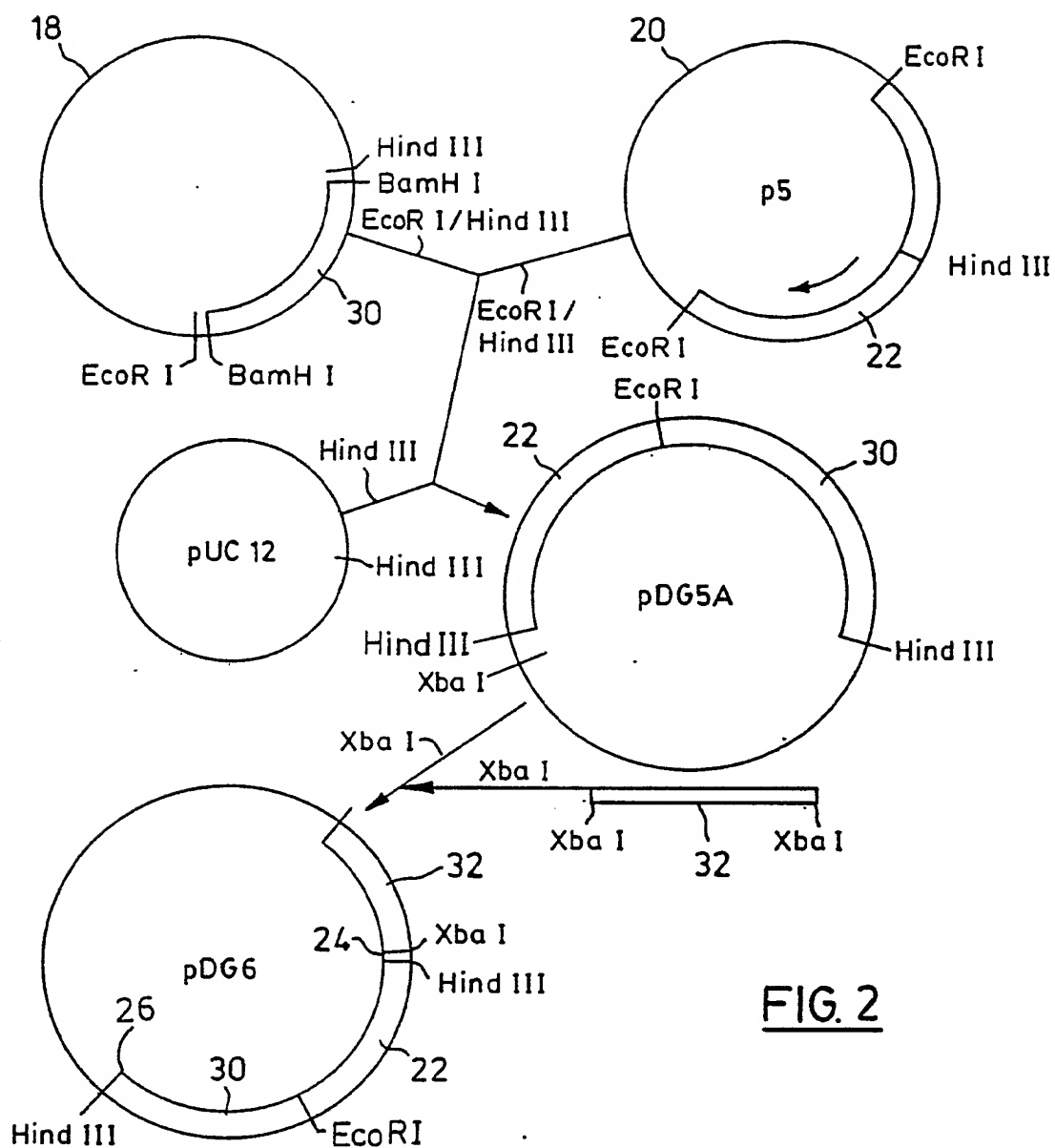


FIG. 2

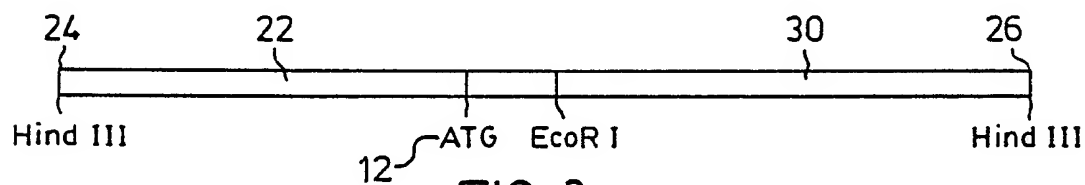


FIG. 3

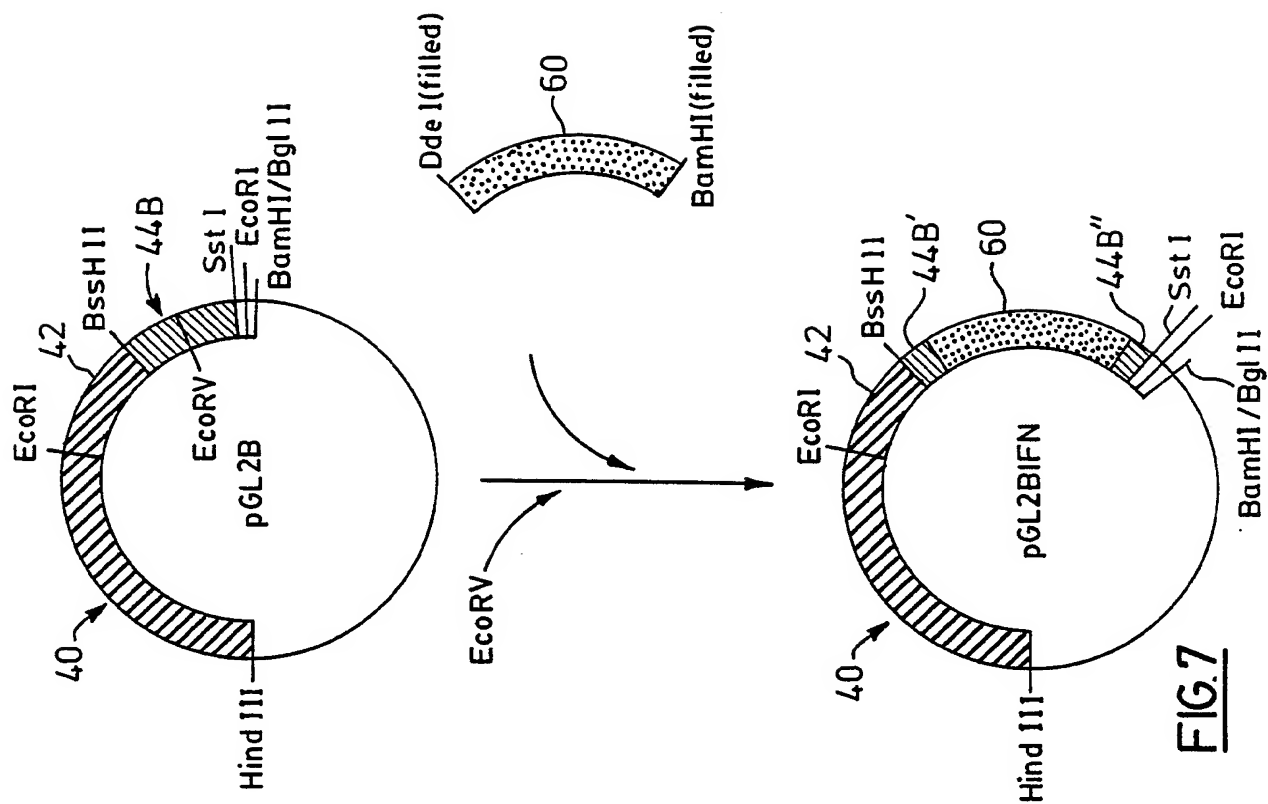


FIG. 7

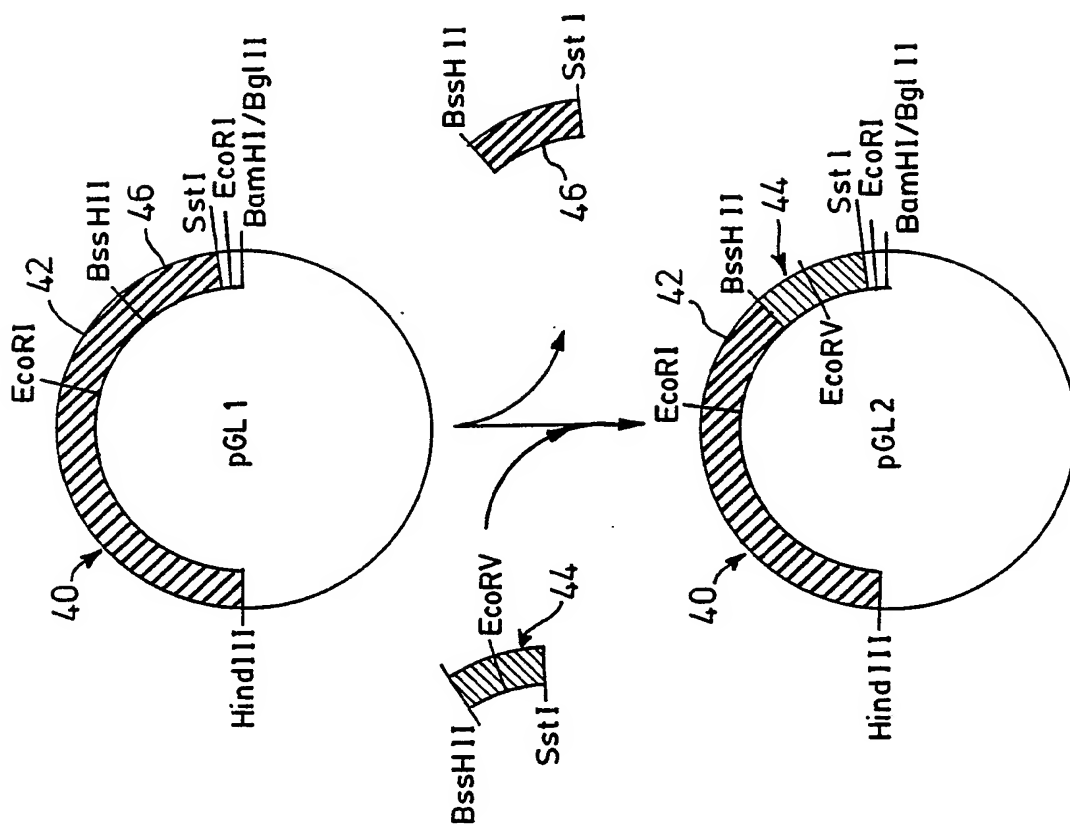
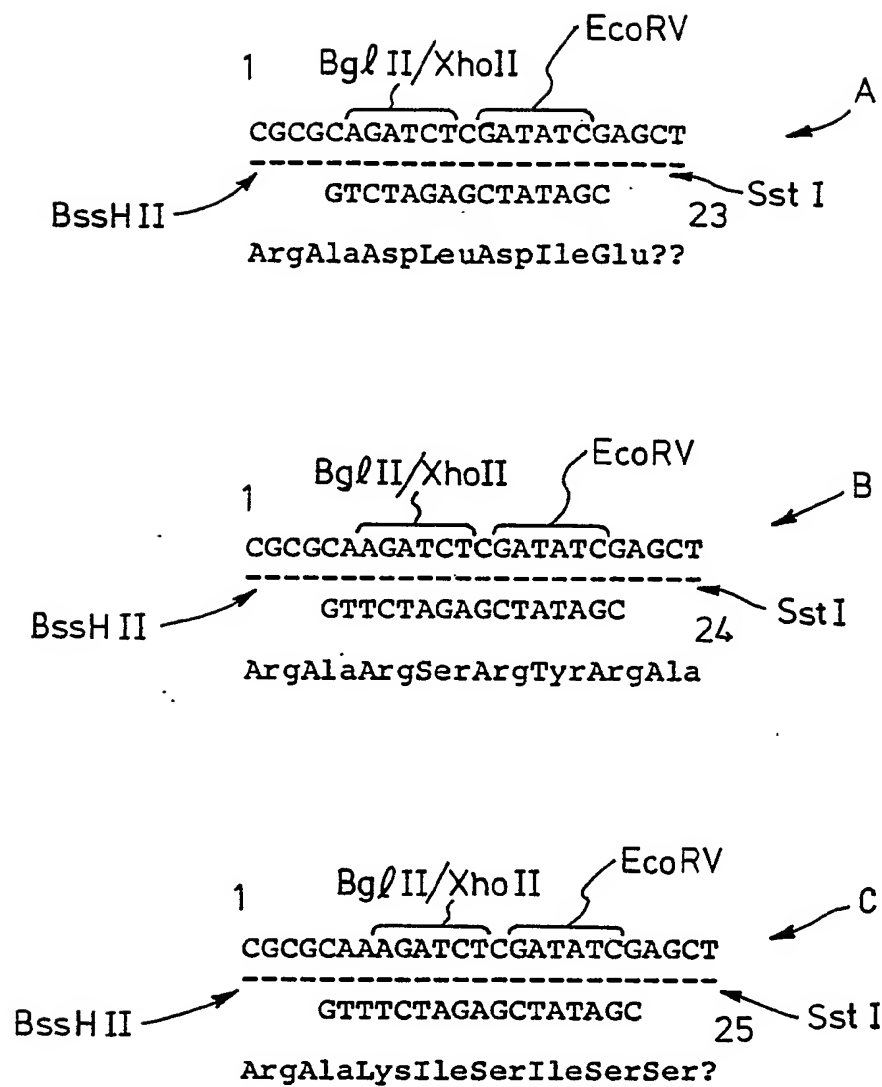


FIG. 4

10/25

FIG.5

11/25

1 EcoRI

GAATTCAAGCTAGATGCTAAGCGATATTGCATGGCAATATGTGTTGATGCATGTGCTTCT
 -----+-----+-----+-----+-----+-----+-----+

TCCTTCAGCTTCCCCTCGTGCAGATGAAGGTTTGGCTATAAATTGAAGTGTTGGTCGGG
 -----+-----+-----+-----+-----+-----+-----+

GTTCCGTGAGGGGCTGAAGTGCTTCCTCCCTTTTAGACGCAACTGAGAGCCTGAGCTTCA
 -----+-----+-----+-----+-----+-----+-----+

TCCCCAGCATCATTACACCTCAGCAATGTCGTTCCGATCTCTACTCGCCCTGAGCGGCCT
 -----+-----+-----+-----+-----+-----+-----+

Met Ser Phe Arg Ser Leu Leu Ala Leu Ser Gly Le

BssHII

CGTCTGCACAGGGTTGGCAAATGTGATTTCCAAGCGCGCG
 -----+-----+-----+-----+-----+-----+-----+

uVal Cys Thr Gly Leu Ala Asn Val Ile Ser Lys Arg Ala

280

FIG. 6

12/25

181 40 49 42

TCCCCAGCATCATTACACCTCAGCAATGTCGTTCCGATCTCTACTCGCCCTGAGCGGCCT
 -----+-----+-----+-----+-----+-----+-----+
 AGGGGTCGTAGTAATGTGGAGTCGTTACAGCAAGGCTAGAGATGAGCGGGACTCGCCGGA
 240
 Met Ser Phe Arg Ser Leu Leu Ala Leu Ser Gly Leu

42 50 52 44B' Ddefilled

241
 CGTCTGCACAGGGTTGGCAAATGTGATTTCCAAGCGCGCAAGATCTCGATTTCAGCTGCAA
 -----+-----+-----+-----+-----+-----+-----+
 GCAGACGTGTCCCAACCGTTTACACTAAAGGTTTCGCGCGTTCTAGAGCTAAGTCGACGTT
 300
 Val Cys Thr Gly Leu Ala Asn Val Ile Ser Lys Arg Ala Arg Ser Arg Phe Ser Cys Lys

301 60

GTCAAGCTGCTCTGTGGGCTGTGATCTGCCTCAAACCCACAGCCTGGGTAGCAGGAGGAC
 -----+-----+-----+-----+-----+-----+-----+
 CAGTTCGACGAGACACCCGACACTAGACGGAGTTTGGGTGTTCGGACCCATCGTCCTCCTG
 360
 Ser Ser Cys Ser Val Gly Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr

361 60 44B" EcoRI

CTTGATGCTC
 -----+-----+-----+-----+-----+-----+-----+
 GAACTACGAG
 AAAACCGGATCATCGAGCTCGAATTC
 -----+-----+-----+-----+-----+-----+-----+
 TTTTGGCCTAGTAGCTCGAGCTTAAG
 Leu Met Leu 1206

FIG. 8

13/25

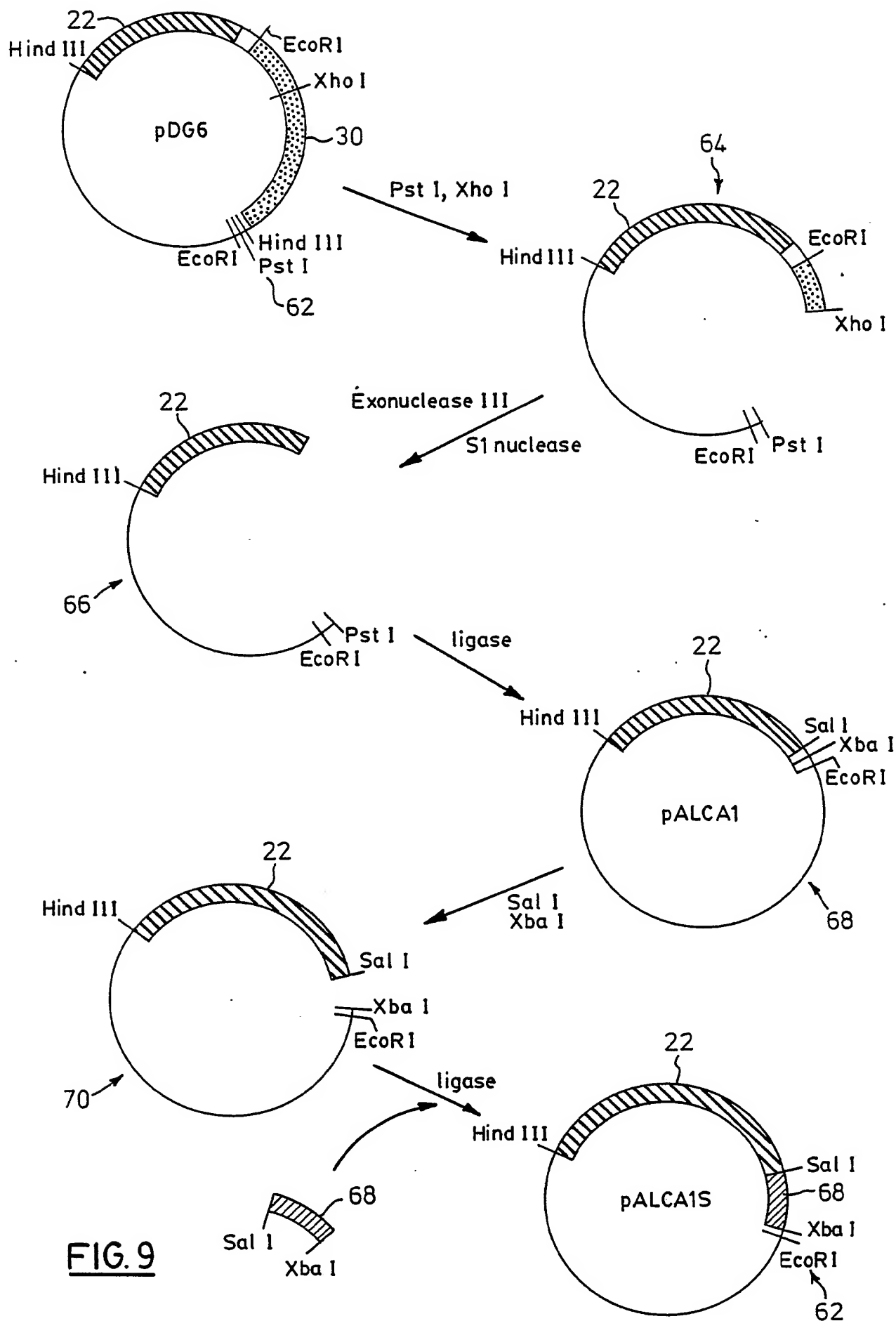


FIG. 9

14/25

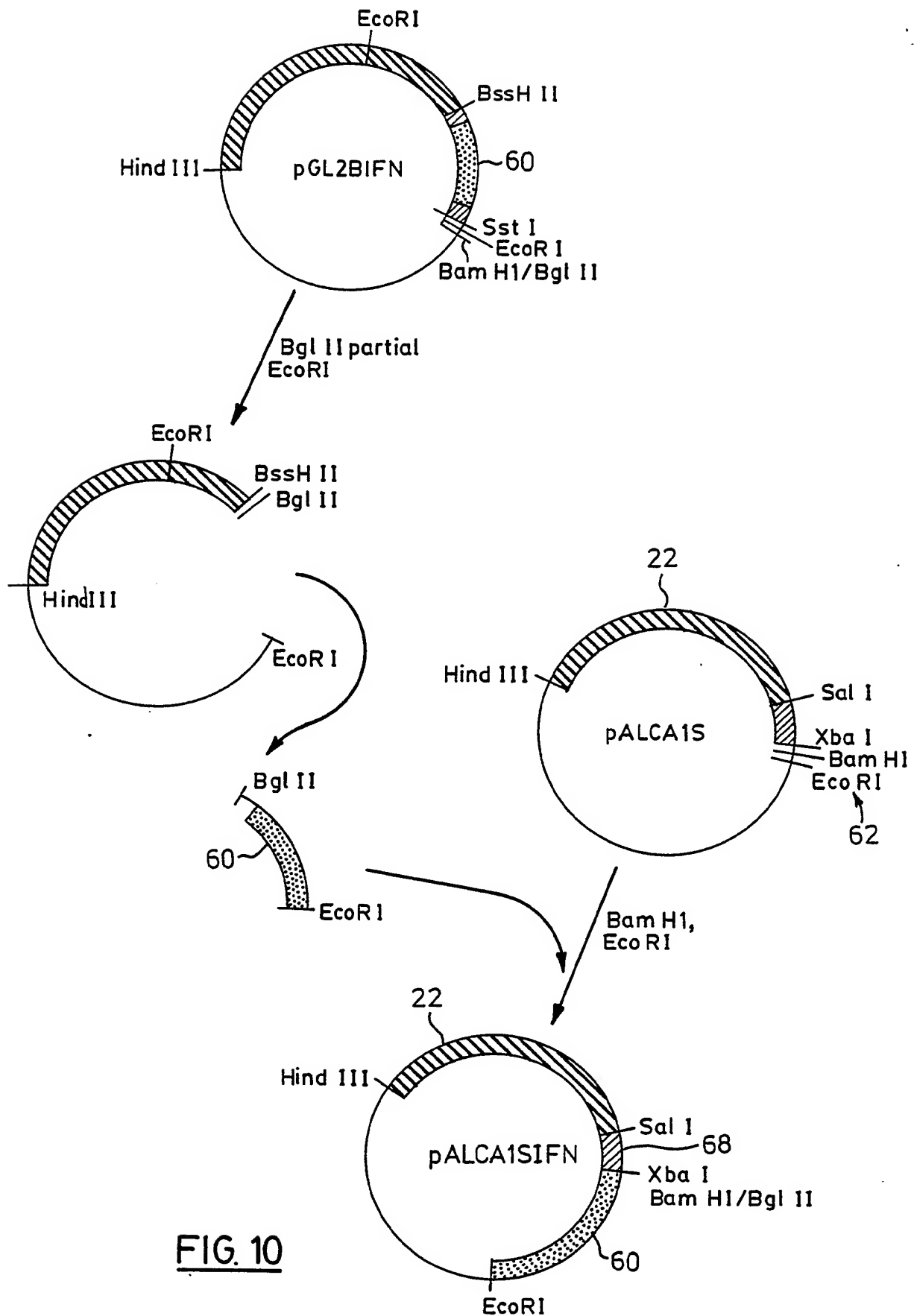


FIG. 10

15/25

1200

GGATACAGTTGGGCATTTCTAGGGCTGAA
-----+-----+-----+
CCTATGTCAACCCGTAAAGATCCCGACTT

TGGGAAGGAGAGAGTTTTGAAATAGGCGTTCCGTTCTGCTTAGGGTATTTGGGAACAATC
-----+-----+-----+-----+-----+
ACCCTTCCTCTCTCAAACTTTATCCGCAAGGCAAGACGAATCCCATAAACCTTGTTAG

AATGTTCAATGTACATTTAATCCACGATTTTATAAAACGTCATCCTTTGCCCTCCCTTCT
-----+-----+-----+-----+-----+
TTACAAGTTACATGTAAATTAGGTGCTAAAATATTTTGCAGTAGGAAACGGGAGGGAAGA

TATTTGCCAATACCAAAAATCTTACTCCAGTGGTTCGGTAATCGCAGAGTTAAATCTGGG
-----+-----+-----+-----+-----+
ATAAACGGTTATGGTTTTTAGAATGAGGTCACCAAGCCATTAGCGTCTCAATTTAGACCC

CTCGGTGGCAGATCTGCGATCGTCCATAACCGTTCAGATGTTGATTGGAAC TGGGTGGGG
-----+-----+-----+-----+-----+
GAGCCACCGTCTAGACGCTAGCAGGTATTGGCAAGTCTACAAC TAACCTTGACCCACCCC

TAGACAGCTCCGAAGACCGAGTGAACGTATACCTAAGACACTTTGACACGGCCGGAACAC
-----+-----+-----+-----+-----+
ATCTGTGCGAGGCTTCTGGCTCACTTGCATATGGATTCTGTGAAACTGTGCCGGCC TTGTG

TGTAAGTCCCTTCGTATTTCTCCGCTGTGTGGAGCTACCATCCAATAACCCCCAGCTGA
-----+-----+-----+-----+-----+
1560

FIG.11 sheet 1

16/25

ACATTCAGGGAAGCATAAAGAGGCGGACACACCTCGATGGTAGGTTATTGGGGGTCGACT

1561

AAAAGCTGATTGTCGATAGTTGTGATAGTTCCCACTTGTCCGTCCGCATCGGCATCCGCA
-----+-----+-----+-----+-----+-----+-----+
TTTTCGACTAACAGCTATCAACACTATCAAGGGTGAACAGGCAGGCGTAGCCGTAGGCGT

GCTCGGGATAGTTCCGACCTAGGATTGGATGCATGCGGAACCGCACgAGGGCGGGGCGGA
-----+-----+-----+-----+-----+-----+-----+
CGAGCCCTATCAAGGCTGGATCCTAACCTACGTACGCCTTGGCGTgcTCCCGCCCCGCT

AATTGACACACCACTCCTCTCCACGCagCCGTTCAGAGGTACGCGTATAGAGCCGTATA
-----+-----+-----+-----+-----+-----+-----+
TTAACTGTGTGGTGAGGAGAGGTGCGTcGGCAAGTTCTCCATGCGCATATCTCGGCATAT

GAGCAGAGACGGAGCACTTCTGGTACTGTCCGCACGGGATGTCCGCACGGAGAGCCACA
-----+-----+-----+-----+-----+-----+-----+
CTCGTCTCTGCCTCGTGAAAGACCATGACAGGCGTGCCCTACAGGCGTGCTCTCGGTGT

AACGAGCGGGGCCCCGTACGTGCTCTCCTACCCCAGGATCGCATCCTCGCATAGCTGAAC
-----+-----+-----+-----+-----+-----+-----+
TTGCTCGCCCCGGGGCATGCACGAGAGGATGGGGTCTAGCGTAGGAGCGTATCGACTTG

ATCTATATAAAGACCCCCAAGGTTCTCAGTCTCACCAACATCATCAACCAACAATCAACA
-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+

FIG.11 sheet 2

1920

17/25

TAGATATATTTCTGGGGGTTCCAAGAGTCAGAGTGGTTGTAGTAGTTGGTTGTTAGTTGT

1921 Sal I 68

GGGTCGACATGTACCGGTTCTCGCCGTCATCTCGGCCTTCCTCGCCACTGCCCTTCGCCA
 -----+-----+-----+-----+-----+-----+-----+
 CCCAGCTGTACATGGCCAAGGAGCGGCAGTAGAGCCGGAAGGAGCGGTGACGGAAGCGGT

Met Tyr Arg Phe Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Phe Ala Lys

Xba I BamHI/BglII fusion 60

AGTCTAGAGGATCTCGATTTCAGCTGCAAGTCAAGCTGCTCTGTGGGCTGTGATCTGCCTC
 -----+-----+-----+-----+-----+-----+-----+
 TCAGATCTCCTAGAGCTAAGTCGACGTTTCAGTTCGACGAGACACCCGACACTAGACGGAG

Ser Arg Gly Ser Arg Phe Ser Cys Lys Ser Ser Cys Ser Val Gly Cys Asp Leu Pro Gln

AAACCCACAGCCTGGGTAGCAGGAGGACCTTGATGCTCCTGGCACAGATGAGGAGAATCT
 -----+-----+-----+-----+-----+-----+-----+
 TTTGGGTGTCGGACCCATCGTCCTCCTGGAACCTACGAGGACCGTGTCTACTCCTCTTAGA

Thr His Ser Leu Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met Arg Arg Ile Ser

CTCTTTTCTCCTGCTTGAAGGACAGACATGACTTTGGATTTCCCCAGGAGGAGTTTGCCA
 -----+-----+-----+-----+-----+-----+-----+
 GAGAAAAGAGGACGAACTTCCTGTCTGTACTGAAACCTAAAGGGGTCCTCCTCAAACCGT

Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Gly Asn

ACCAGTTCCAAAAGGCTGAAACCATCCCTGTCCTCCATGAGATGATCCAGCAGATCTTCA
 -----+-----+-----+-----+-----+-----+-----+
 TGGTCAAGGTTTTCCGACTTTGGTAGGGACAGGAGGTACTCTACTAGGTCGTCTAGAAGT

Gln Phe Gln Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile Phe Asn 2220

FIG.11 sheet 3

18/25

2221

ATCTCTTCAGCACAAAGGACTCATCTGCTGCTTGGGATGAGACCCTCCTAGACAAATTCT
-----+-----+-----+-----+-----+-----+-----+
TAGAGAAGTCGTGTTTCCTGAGTAGACGACGAACCCTACTCTGGGAGGATCTGTTTAAGA

Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr

ACACTGAACTCTACCAGCAGCTGAATGACCTGGAAGCCTGTGTGATACAGGGGGTGGGGG
-----+-----+-----+-----+-----+-----+-----+
TGTGACTTGAGATGGTCGTCGACTTACTGGACCTTCGGACACACTATGTCCCCCACCCCC

Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Gly Val Gly Val

TGACAGAGACTCCCCTGATGAAGGAGGACTCCATTCTGGCTGTGAGGAAATACTTCCAAA
-----+-----+-----+-----+-----+-----+-----+
ACTGTCTCTGAGGGGACTACTTCCTCCTGAGGTAAGACCGACACTCCTTTATGAAGGTTT

Thr Glu Thr Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg

GAATCACTCTCTATCTGAAAGAGAAGAAATACAGCCCTTGTGCCTGGGAGGTTGTCAGAG
-----+-----+-----+-----+-----+-----+-----+
CTTAGTGAGAGATAGACTTTCTCTTCTTTATGTCGGGAACACGGACCCTCCAACAGTCTC

Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val Arg Ala

CAGAAATCATGAGATCTTTTCTTTGTCAACAAACTTGCAAGAAAGTTTAAGAAGTAAGG
-----+-----+-----+-----+-----+-----+-----+
GTCTTTAGTACTCTAGAAAAAGAAACAGTTGTTTGAACGTTCTTTCAAATTCTTCATTCC

Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser Leu Arg Ser Lys Glu

60

2520

FIG.11 sheet 4

19/25

2521

AATGAAAAC TGGTTCAACATGGAAATGATTTTCATTGATTCGTATGCCAGCTCACCTTTT
-----+-----+-----+-----+-----+-----+-----+
TTACTTTTGACCAAGTTGTACCTTTACTAAAAGTAACTAAGCATACGGTCGAGTGGA AAA

End

TATGATCTGCCATTTCAAAGACTCATGTTTCTGCTATGACCATGACACGATTTAAATCTT
-----+-----+-----+-----+-----+-----+-----+
ATACTAGACGGTAAAGTTTCTGAGTACAAAGACGATACTGGTACTGTGCTAAATTTAGAA

TTCAAATGTTTTTAGGAGTATTAATCAACATTGTATTCAGCTCTTAAGGCACTAGTCCCT
-----+-----+-----+-----+-----+-----+-----+
AAGTTTACAAAAATCCTCATAATTAGTTGTAACATAAGTCGAGAATTCCGTGATCAGGGA

TACAGAGGACCATGCTGACTGATCCATTATCTATTTAAATATTTTTAAATATTATTTAT
-----+-----+-----+-----+-----+-----+-----+
ATGTCTCCTGGTACGACTGACTAGGTAATAGATAAATTTATAAAAATTTTATAATAAATA

TTAACTATTTATAAAACAAC TTATTTTGTTCATATTATGTCATGTGCACCTTTGCACAG
-----+-----+-----+-----+-----+-----+-----+
AATTGATAAATATTTTGTGTAATAAAACAAGTATAATACAGTACACGTGGAAACGTGTC

TGGTTAATGTAATAAAATATGTTCTTTGTATTTGGTAAAAAAAAAAAAAAAAAAAAA
-----+-----+-----+-----+-----+-----+-----+
ACCAATTACATTATTTTATACAAGAAACATAAACCATTTTTTTTTTTTTTTTTTTTTT

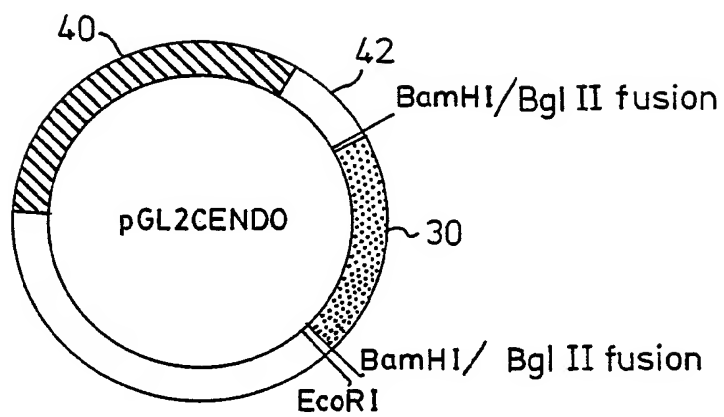
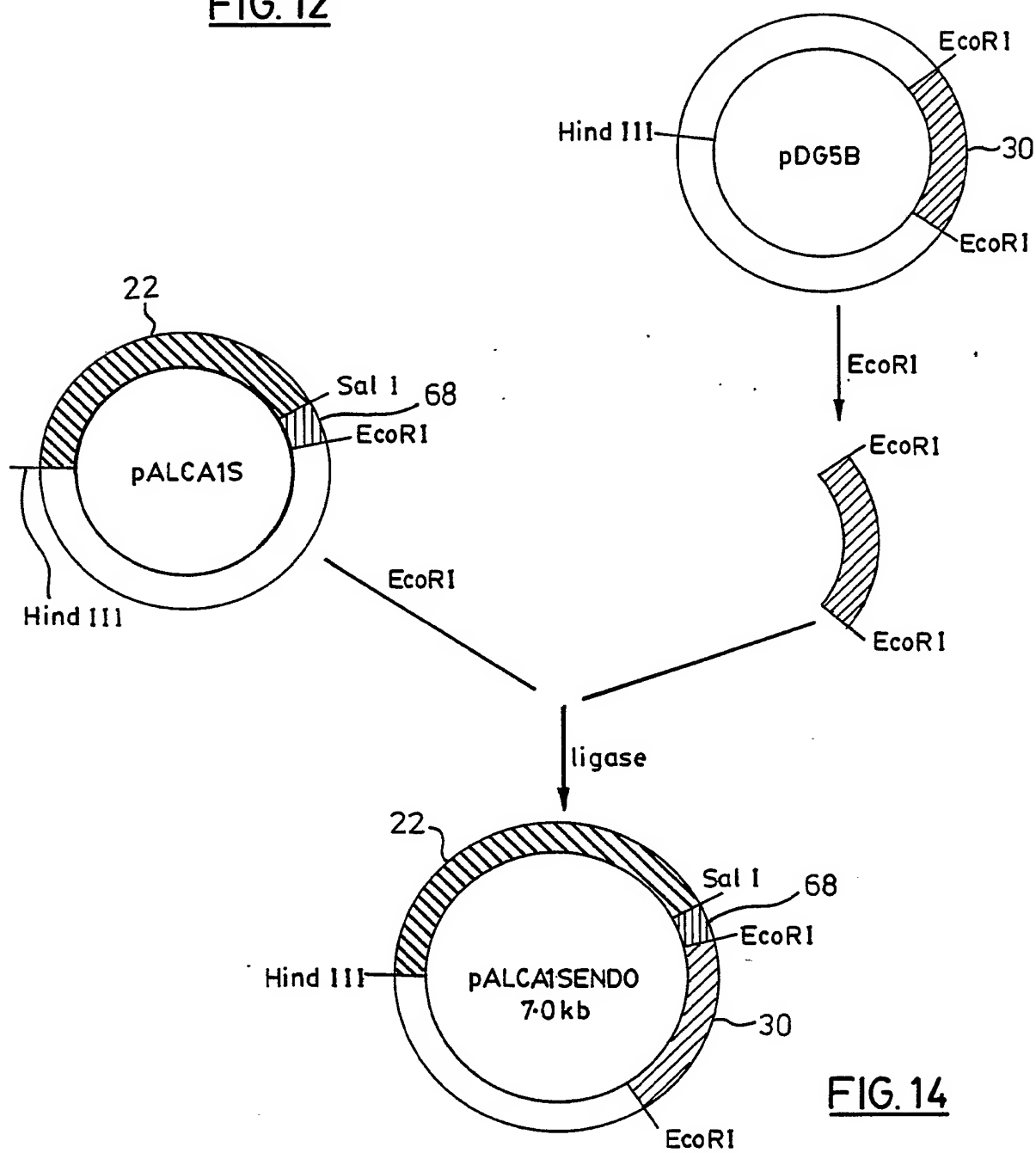
EcoRI

AAAAAAAAAAAAACCGGATCATCGAGCTCGAATTC
-----+-----+-----+-----+-----+-----+-----+
TTTTTTTTTTTTTGGCCTAGTAGCTCGAGCTTAAG

2914

FIG. 11 sheet 5

20/25

FIG. 12FIG. 14

21/25

1 EcoRI
 GAATTCAAGCTAGATGCTAAGCGATATTGCATGGCAATATGTGTTGATGCATGTGCTTCT
 -----+-----+-----+-----+-----+-----+-----+
 CTTAAGTTCGATCTACGATTCGCTATAACGTACCGTTATACACAACCTACGTACACGAAGA

TCCTTCAGCTTCCCCCTCGTGCAGATGAAGGTTTGGCTATAAATTGAAGTGGTTGGTCGGG
 -----+-----+-----+-----+-----+-----+-----+
 AGGAAGTCGAAGGGGAGCACGTCTACTTCCAAACCGATATTTAACTTCACCAACCAGCCC

GTTCCGTGAGGGGCTGAAGTGCTTCCTCCCTTTTAGACGCAACTGAGAGCCTGAGCTTCA
 -----+-----+-----+-----+-----+-----+-----+
 CAAGGCACTCCCCGACTTCACGAAGGAGGGAAAATCTGCGTTGACTCTCGGACTCGAAGT

TCCCCAGCATCATTACACCTCAGCAATGTCGTTCCGATCTCTACTCGCCCTGAGCGGCCT
 -----+-----+-----+-----+-----+-----+-----+
 AGGGGTCGTAGTAATGTGGAGTCGTTACAGCAAGGCTAGAGATGAGCGGGACTCGCCGGA

Met Ser Phe Arg Ser Leu Leu Ala Leu Ser Gly Leu 240

BamHI/BglII fusion
 241 42 30 EcoRI
 CGTCTGCACAGGGTTGGCAAATGTGATTTCCAAGCGCGCAAAGATCCAG...GAATTC
 -----+-----+-----+-----+-----+-----+-----+
 GCAGACGTGTCCAACCGTTTACACTAAAGGTTTCGCGCGTTTCTAGGTC...CTTAAG

Val Cys Thr Gly Leu Ala Asn Val Ile Ser Lys Arg Ala Lys Ile Gln

FIG.13

22/25

1381

CTCGGTGGCAGATCTGCGATCGTCCATAACCGTTCAGATGTTGATTGGAAC TGGGTGGG
-----+-----+-----+-----+-----+-----+
GAGCCACCGTCTAGACGCTAGCAGGTATTGGCAAGTCTACAAC TAACCTTGACCCACCCC

TAGACAGCTCCGAAGACCGAGTGAACGTATACCTAAGACACTTTGACACGCGCGGAACAC
-----+-----+-----+-----+-----+-----+
ATCTGTCGAGGCTTCTGGCTCAGTTGCATATGGATTCTGTGAAACTGTGCCGGCCTTGTC

TGTAAGTCCCTTCGTATTTCTCCGCCTGTGTGGAGCTACCATCCAATAACCCCCAGCTGA
-----+-----+-----+-----+-----+-----+
ACATTCAAGGAAGCATAAAGAGGCGGACACACCTCGATGGTAGGTTATTGGGGGTGACT

AAAAGCTGATTGTCGATAGTTGTGATAGTTCCCACTTGTCCGTCCGCATCGGCATCCGCA
-----+-----+-----+-----+-----+-----+
TTTTCGACTAACAGCTATCAACACTATCAAGGGTGAACAGGCAGGCGTAGCCGTAGGCGT

GCTCGGGATAGTTCCGACCTAGGATTGGATGCATGCGGAACCGCACgAGGGCGGGGCGGA
-----+-----+-----+-----+-----+-----+
CGAGCCCTATCAAGGCTGGATCCTAACCTACGTACGCCTTGGCGTgcTCCCGCCCCGCCT

AATTGACACACCACTCCTCTCCACGCAgCCGTTCAAGAGGTACGCGTATAGAGCCGTATA
-----+-----+-----+-----+-----+-----+
TTAACTGTGTGGTGAGGAGAGGTGCGTcGGCAAGTTCTCCATGCGCATATCTCGGCATAT

1740

FIG.15 sheet 1

23/25

1741

GAGCAGAGACGGAGCACTTTCTGGTACTGTCCGCACGGGATGTCCGCACGGAGAGCCACA
 -----+-----+-----+-----+-----+-----+-----+
 CTCGTCTCTGCCTCGTGAAAGACCATGACAGGCGTGCCCTACAGGCGTGCCCTCTCGGTGT

AACGAGCGGGGCCCCGTACGTGCTCTCCTACCCCAGGATCGCATCCTCGCATAGCTGAAC
 -----+-----+-----+-----+-----+-----+-----+
 TTGCTCGCCCCGGGGCATGCACGAGAGGATGGGGTCCTAGCGTAGGAGCGTATCGACTTG

ATCTATATAAAGACCCCCAAGGTTCTCAGTCTCACCAACATCATCAACCAACAATCAACA
 -----+-----+-----+-----+-----+-----+-----+
 TAGATATATTTCTGGGGGTTCCAAGAGTCAGAGTGGTTGTAGTAGTTGGTTGTTAGTTGT

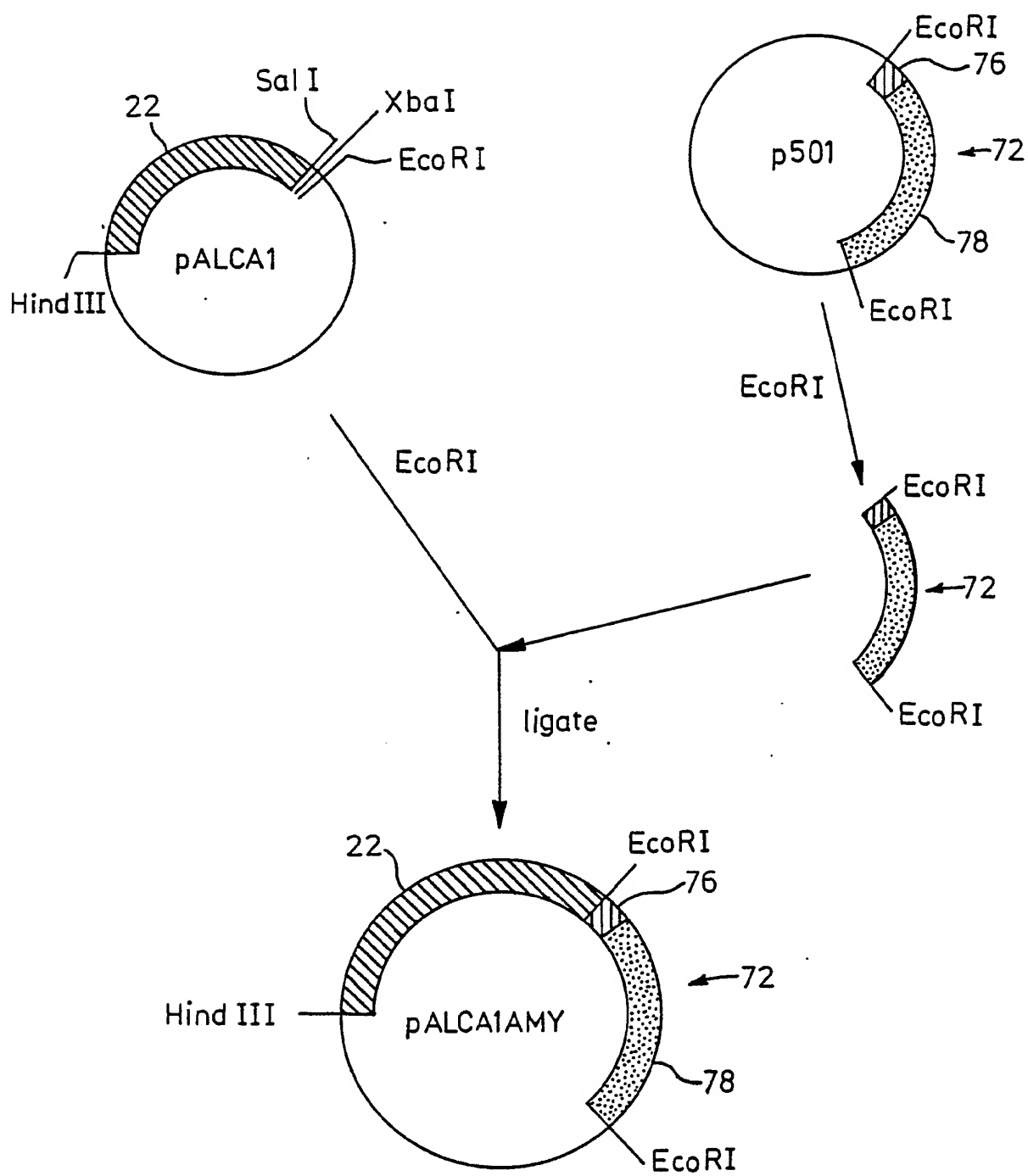
Sal I 68
 GGGTCGACATGTACCGGTTCTCGCCGTCATCTCGGCCTTCCTCGCCACTGCCTTCGCCA
 -----+-----+-----+-----+-----+-----+-----+
 CCCAGCTGTACATGGCCAAGGAGCGGCAGTAGAGCCGGAAGGAGCGGTGACGGAAGCGGT
 1980
 Met Tyr Arg Phe Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Phe Ala Lys

EcoRI 30
 1981
 AGTCTAGAGGATCCCCGGGCGAGCTCGAATTCCTCGGGGATCCAG
 -----+-----+-----+-----+-----+-----+-----+
 TCAGATCTCCTAGGGGCCCGCTCGAGCTTAAGGGCCCCCTAGGTC
 Ser Arg Gly Ser Pro Gly Glu Leu Glu Phe Pro Gly Ile Gln endoglucanase coding

EcoRI
 CCGGGGCGAGCTCGAATTC
 +-----+-----+-----+
 GGGCCCCGCTCGAGCTTAAC

FIG.15 sheet 2

24/25

FIG. 16

25/25

361

AGGTTCTCAGTCTCACCAACATCATCAACCAACAATCAACAGGGTTCGACTCTAGAGGATC
 -----+-----+-----+-----+-----+-----+-----+-----+
 TCCAAGAGTCAGAGTGGTTGTAGTAGTTGGTTGTTAGTTGTCCCAGCTGAGATCTCCTAG

22 EcoRI
 -----+-----+-----+-----+-----+-----+-----+-----+
 CCGGGGCGAGCTCGAATTCCCGGATCCGTCGACCTGCAGGGGGGGGGGGGGTCTTCTCC
 -----+-----+-----+-----+-----+-----+-----+-----+
 GGGCCCGCTCGAGCTTAAGGGGCCCTAGGCAGCTGGACGTCCCCCCCCCCCCCAGAAGAGG

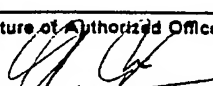
76
 -----+-----+-----+-----+-----+-----+-----+-----+
 ACCGTCCTCTTGCAGAGCACACAGAGCTGAAGACGATGGCGAACAAACATCTGTCCCT
 -----+-----+-----+-----+-----+-----+-----+-----+
 TGGCAGGAGAACGTCTCGTGTGTGTCGACTTCTGCTACCGCTTGTTTGTAGACAGGGA
 MetAlaAsnLysHisLeuSerLeu

76 78
 -----+-----+-----+-----+-----+-----+-----+-----+
 CTCGCTCTTCTCGTCCTCCTTGGCCTGTGGCATCTCTAGCTTCCGGCCAA
 -----+-----+-----+-----+-----+-----+-----+-----+
 GAGCGAGAAGGAGCAGGAGGAACCGGACAGCCGTAGAGATCGAAGGCCGGTT 592
 SerLeuPheLeuValLeuLeuGlyLeuSerAlaSerLeuAlaSerGlyGln 592

FIG.17

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 86/00209

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC ⁴ : C 12 N 15/52; C 12 N 15/67; C 12 N 15/80; C 12 N 15/28; C 12 N 15/26		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
IPC ⁴	C 12 N	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched *		
III. DOCUMENTS CONSIDERED TO BE RELEVANT *		
Category *	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	EP, A, 0126206 (CETUS) 28 November 1984 see page 5, lines 24-28; page 10, line 11 - page 12, line 4; page 12, lines 13-16; page 15, lines 3-9; page 21, table I	1-3,7-15
Y	--	4-6,16-88
X	EP, A, 0099226 (ST. LOUIS UNIVERSITY) 25 January 1984, see page 4, lines 15-22; page 6, line 34 - page 7, line 31; page 15, lines 29-32; page 17, line 3 - page 18, line 7	1-3,7-15
Y	--	4-6,16-88
X	Chemical Abstracts, vol. 98, no. 25, 20 June 1983 (Columbus, Ohio, US) D.J. Ballande et al.: "Transformation of Aspergillus nidulans by the orotidine 5'-phosphate decarboxylase gene of Neurospora crassa", see page 150, abstract no. 210814t & Biochem. Biophys. Res. Commun. 1983, 112(1), 284-9	1,2,10,56- 58,60,67-70
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>* Special categories of cited documents: ¹⁰</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
24th July 1986		08 SEP 1986
International Searching Authority		Signature of Authorized Officer
EUROPEAN PATENT OFFICE		 L. ROSSI

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X	Chemical Abstracts, vol. 103, 29 July 1985 (Columbus, Ohio, US). C.H. Doy et al.: "Genomic clones of <i>Aspergillus nidulans</i> containing <i>alcA</i> , the structural gene for alcohol dehydrogenase and <i>alcR</i> , a regulatory gene for ethanol metabolism", see page 153, abstract no. 17752q & DNA 1985, 4(2), 105-14	1-5, 10-15
A	--	15-29, 34-41 48-64, 68-88
X	Chemical Abstracts, vol. 101, no. 17, 22 October 1984 (Columbus, Ohio, US) E. Boel et al.: "Two different types of intervening sequences in the glucoamylase gene from <i>Aspergillus niger</i> ", see page 158, abstract no. 145009y & EMBO J. 1984, 3(7), 1581-5	1-3, 7-15
A	--	18-22, 25-72
X	Gene, volume 33, 1985; Elsevier Science Publ.; R.A. Lockington et al.: "Cloning and characterization of the ethanol utilization regulon in <i>Aspergillus nidulans</i> ", pages 137-149	1-5, 10-15
A	--	15-29, 34-41, 48-64, 68-88
P, X	Chemical Abstracts, vol. 104, 1986 (Columbus, Ohio, US) C.A.M.J.J. Van den Hondel et al.: "Development of a system for analysis of regulation signals in <i>Aspergillus</i> ", see page 167, abstract no. 103393a & Symp. Mol. Cell. Biol., New Ser. 1985, 34 (Mol. Genet. Filamentous Fungi), 29-38	1-4, 10-14, 56 -58, 60, 68-70
P, X	Gene, vol. 37, 1985, Elsevier Science Publishers; F.P. Buxton et al.: "Transformation of <i>Aspergillus niger</i> using the <i>argB</i> gene of <i>Aspergillus nidulans</i> ", pages 207-214, see summary	1-4, 10, 56, 57 59, 68, 69, 71
P, X	Chemical Abstracts, vol. 98, no. 19, 9 May 1983 (Columbus, Ohio, US) J.A. Pateman et al.: "Regulation of alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (AldDH) in <i>Aspergillus nidulans</i> ", see page 247,	./.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
	abstract no. 157629h & Proc. R. Soc. London, (Ser) B 1983, 217(1208), 243-64 --	72-77,83-88
Y	EP, A, 0136907 (GENENTECH) 10 April 1985 see page 2, line 31 - page 3, line 5 --	72-77, 83-88
A	Nature, vol. 308, no. 5960, April 1984 (Reading, Berks, GB) S.J. Rothstein et al.: "Secretion of a wheat α -amylase expressed in yeast", pages 662-665, see the whole document --	23-24,61,62
A	Gene, vol. 17, 1982, Elsevier Biomedical Press; D.J. Whittle et al.: "Molecular cloning of a cellulomonas fimi cellulase gene in escherichia coli", pages 139-145, see summary --	16,17,43,44, 52,53,63,66
A	Gene, vol. 12, 1980, Elsevier Biomedical Press; K. Talmadge et al.: "Construc- tion of plasmid vectors with unique PstI cloning sites in a signal sequence coding region", pages 235-241 see summary --	25,36
A	EP, A, 0127304 (GENENTECH) 5 December 1984 see page 11, lines 20-32; claims 1,6 --	45,46,54,55, 64,65
A	Chemical Abstracts, vol. 100, no. 5, 30 January 1984 (Columbus, Ohio, US) J.A. Pateman et al.: "Genes for alco- hol utilization in the lower eukaryote Aspergillus nidulans", see page 108, abstract no. 30390c & Manipulation Expression Genes Eukaryotes Proc. Int. Conf., Int. Congr. Biochem., 12th 1982 (Pub. 1983) 171-8 --	1-6,10-29,34- 41,48-64,68- 88
E	EP, A, 0181213 (ANTIBIOTICOS) 14 May 1986 see claims --	1,2,10,11,56 67,68
E	EP, A, 0173378 (UNILEVER) 5 March 1986 see claims 40-44 -----	1,2,10,11,56 -57,59,60,67 -69,71

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON

INTERNATIONAL APPLICATION NO. PCT/GB 86/00209 (SA 12930)

This Annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 18/08/86

The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0126206	28/11/84	WO-A- 8402921 AU-A- 2499384 JP-A- 60070079	02/08/84 15/08/84 20/04/85
EP-A- 0099226	25/01/84	JP-A- 59042891 CA-A- 1207687 US-A- 4486533	09/03/84 15/07/86 04/12/84
EP-A- 0136907	10/04/85	AU-A- 3375984	18/04/85
EP-A- 0127304	05/12/84	AU-A- 2721884 JP-A- 60041488	01/11/84 05/03/85
EP-A- 0181213	14/05/86	None	
EP-A- 0173378	05/03/86	JP-A- 61092569	10/05/86

For more details about this annex :
see Official Journal of the European Patent Office, No. 12/82